Universidade Federal do Amazonas
Instituto de Computação
Programa de Pós-Graduação em Informática

UM MODELO DE REPUTAÇÃO COM CLASSIFICAÇÃO VIA AGRUPAMENTO PARA DETECÇÃO DE NÓS EGOÍSTAS EM REDES OPORTUNISTAS

Manaus

Maio de 2018

DIOGO SOARES MOREIRA

UM MODELO DE REPUTAÇÃO COM CLASSIFICAÇÃO VIA AGRUPAMENTO PARA DETECÇÃO DE NÓS EGOÍSTAS EM REDES OPORTUNISTAS

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito parcial para a obtenção do grau de Mestre em Informática.

Orientador: Dr.-Ing Edjair de Souza Mota

Manaus

Maio de 2018

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Soares Moreira, Diogo
S676u Um Modelo de Reputação com Classificação via Agrupamento para Detecção de Nós Egoístas em Redes Oportunistas / Diogo Soares Moreira. 2018
93 f.: il. color; 31 cm.

Orientador: Edjair de Souza Mota
Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas.

1. Detecção de Egoísmo. 2. Redes Tolerantes a Atrasos e Desconexões. 3. Redes Oportunistas Móveis. 4. Modelos de Reputação. I. Mota, Edjair de Souza II. Universidade Federal do Amazonas III. Título

PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UFAM

FOLHA DE APROVAÇÃO

"Um Modelo de Reputação com Classificação via Agrupamento para Deteção de Nós Egoístas em Redes Oportunistas"

DIOGO SOARES MOREIRA

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof. Edjair Souza Mota - PRESIDENTE

Prof. Alexandre Passito de Queiroz - MEMBRO EXTERNO

Prof. Leandro Silva Galvão de Carvalho - MEMBRO EXTERNO

Prof. Celso Barbosa Carvalho - MEMBRO EXTERNO

Manaus, 11 de Maio de 2018

Dedico este trabalho a minha mãe Zélia e minha amada, Jessyca.

Agradecimentos

Agradeço primeiramente a Deus, por toda força e iluminação nos momentos mais necessários desta vida e permitindo que eu alcance tais objetivos que jamais sonharia em alcançar.

Agradecimento especial ao meu orientador, professor Edjair, que mesmo nos momentos em que não estive tão engajado com o mestrado, me apoiou e me incentivou a finaliza-lo, diversas vezes, sendo sempre paciente e determinado com a proposta deste trabalho.

Também agradeço a meus familiares, em especial à minha mãe que muito me incentivou nos estudos e até mesmo neste mestrado quando não parecia haver mais forças para terminar.

A Jessyca, minha amada, que foi paciente, amigável e sempre tentou me apoiar e me ajudou por diversas vezes na escrita deste trabalho.

Ademais, agradeço a amigos de longa data da UFAM e da UEA que sempre me incentivaram e deram a maior força dizendo que iria vencer esta etapa da vida, em especial aos amigos José Raulinho, Jessika Batista, Jeane Galves, Angela Emi, Pamela Nunes e Silmara Dias por todas as palavras ditas nos melhores e, principalmente, nos momentos que eu dizia "Dessa vez não vou conseguir, pessoal!".

A todos vocês, meu muito obrigado, essa é pra vocês!

"Yeah science!"
(Jesse Pinkman)

Resumo

O avanço das tecnologias em infraestrutura de redes fez emergir uma nova gama de aplicações que se utilizam de múltiplos saltos, tais como as redes de sensores e as redes em malha sem fio. O conceito de redes tolerantes a atrasos e desconexões surgiu como uma solução para possibilitar a comunicação em cenários nos quais a comunicação é intermitente. Todavia, uma premissa básica para o bom desempenho dessas redes é a colaboração dos nós durante a comunicação. Essa colaboração é uma questão fundamental para o fluxo de dados e o desempenho global. Entretanto, devido a fatores como restrições de recursos próprios (por exemplo, *buffer* e energia), os nós eventualmente agem de maneira egoísta, deixando de colaborar com o fluxo de dados na rede. Portanto, é fundamental que exista um mecanismo distribuído que possa mensurar o grau de colaboração dos membros da rede a fim de melhorar a entrega das mensagens. Este trabalho propõe um modelo de detecção de egoísmo utilizando mecanismos de reputação, que possa qualificar outros membros da rede de acordo com seu grau de participação no fluxo de dados, podendo, desta forma, identificar nós egoístas e nós cooperativos. O ranqueamento é feito através de um método numérico executado durante os contatos. Uma vez qualificados, os membros da rede são classificados através de uma técnica de agrupamento, diferentemente de outros trabalhos encontrados na literatura. Os resultados experimentais, obtidos no ambiente de simulação The ONE, demonstram que o modelo aqui proposto e implementado é promissor e a técnica de agrupamento pode ser aplicada sem perda de confiabilidade, além se ser muito preciso quando a taxa de nós egoístas na rede aumenta.

Palavras-chave: Detecção de Egoísmo, Redes Tolerantes a Atrasos e Desconexões, Redes Oportunistas Móveis, Modelos de Reputação.

Abstract

The advance of networking infrastructure technologies has spawned a new range of multi-hop applications, such as sensor networks and wireless mesh networks. The concept of delay and disruption-tolerant networks has emerged as a solution to enable communication in scenarios in which communication is intermittent. However, a basic premise for the good performance of these networks is the collaboration of the nodes during the communication. Such collaboration is a critical issue to the data flow and overall performance. However, due to factors resources constraints (e.g., buffer and energy), the nodes can act selfishly, and stop collaborating with the data flow in the network. The existence of a distributed mechanism to measure the collaboration degree of the network nodes is, therefore, vital to improve the delivery of messages. This work proposes a model of egoism detection using reputation mechanisms, which can qualify other members of the network according to their participation degree in the data flow to identify both selfish and cooperative nodes. The proposed model makes use of a numerical-based ranking method through executed during the contacts. Once qualified, a grouping technique classifies the network nodes, which differs from other works found in the literature. The experimental results, got in the simulation environment The ONE, show that the model proposed and implemented here is promising and the clustering technique can be applied without loss of reliability, and being very precise when the rate of selfish nodes in the network increases.

Keywords: Selfishness Detection, Delay Tolerant Networks, Opportunistic Mobile Networks, Reputation System.

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

Lista de Abreviaturas

ACK Acknowledgement
CORE A Collaborative Reputation Mechanism to Enforce Node Cooperation in Mobile Ad hoc Networks
CONFIDANT Cooperation Of Nodes: Fairness In Dynamic Ad-hoc NeTworks
DTN Delay Tolerant Networks
IP Internet Protocol
MANET Mobile Ad Hoc Network
OCEAN Observation-based Cooperation Enforcement in Ad hoc Networks
ONE Opportunistic Network Environment
OSI Open System Interconnection
PDU Protocol Data Units
P2P Peer to Peer
SSAR Social Selfishness Aware Routing
TCP Transmission Control Protocol
TPR True Positive Rate

Sumário

Capítulo 1

Introdução

O avanço recente na tecnologia de telecomunicações e de informação culminou numa vasta gama de dispositivos interconectados em uma rede infraestruturada, a *Internet*. Na infraestrutura da *Internet*, o compartilhamento de informações de modo confiável é uma importante parte do processo de comunicação entre os dispositivos tais como computadores, aparelhos celulares, entre outros. Isso foi possível devido às características internas da pilha de protocolos TCP/IP, um conjunto de protocolos que permite a comunicação entre dispositivos de modo confiável. Contudo, há certos cenários nos quais as premissas para bom funcionamento da pilha de protocolos TCP/IP não são atendidas tais como: redes rurais, redes de sensores, redes *ad hoc* móveis e redes oportunistas, tornando necessário o desenvolvimento de um novo conjunto de protocolos capazes de atender a estes ambientes desafiadores.

Redes tolerantes a atrasos e desconexões (DTN – *Delay Tolerant Networks*) (Fall, 2003) são conjuntos de protocolos capazes de prover comunicação em ambientes definidos pela falta de infraestrutura fixa de rede e altas taxas de atraso. Assim, DTN surgiram como opção para cenários de conectividade esporádica tais como ambientes militares, rurais e, até mesmo, urbanos como redes de celulares interconectadas ou redes de veículos. Uma DTN consiste de um conjunto de dispositivos, os nós, que se comunicam por mensagens de forma assíncrona, isto é, se um nó X deseja repassar uma mensagem para o nó Y tal que não haja um caminho fim a fim entre ambos, então X pode encaminhar a mensagem através do contato oportunista com outros nós que compõem essa rede, pois a cópia da mensagem durante o repasse, aumenta a probabilidade de entrega para o destinatário final. Entretanto, comunicações sem infraestrutura fixa de rede estão sujeitas a problemas como longos atrasos e erros.

O tipo de redes no qual o repasse de dados é feito de maneira oportunista utilizando-se a mobilidade humana é chamado de redes oportunistas (*Opportunistic*

Networks), que é uma subcategoria de DTN.

Redes sem essa infraestrutura dependem, para um melhor desempenho de transferência de mensagens, que os nós repassem adiante tráfego de mensagens não relacionados a seu próprio uso, de maneira oportunista, isto é, que aceitem carregar consigo mensagens destinadas a outros nós, elevando a probabilidade de entrega ao aumentar a quantidade de réplicas disponíveis entre os nós da rede. Assim, a comunicação é dependente de esquemas cooperativos durante a transferência de mensagens para garantia de um desempenho mais satisfatório. Contudo, em alguns ambientes reais, o modelo de transmissão das redes DTN apresenta algumas limitações. Um destes está relacionado à participação dos nós da rede, pois, estes podem, dependendo de algumas circunstâncias atuar de modo chamado egoísta, isto é, nós da rede que não cooperam com a comunicação ao não realizar o repasse de mensagens de outros nós. Isto acontece pois os nós podem desejar economizar alguns de seus recursos (energia e *buffer*, por exemplo) para que não sejam utilizados ao máximo, em razão da condução de tráfego destinado a outros membros da rede.

Os nós da rede que rejeitam carregar mensagens consigo de modo intermediário são chamados de nós egoístas. Assim, esses nós almejam se beneficiar da rede para que os outros nós vizinhos repassem suas mensagens, ao mesmo tempo que, seus recursos são maximizados devido ao egoísmo.

O impacto do egoísmo tem sido estudo previamente em (Agrawal, 2005) utilizando roteamento epidêmico. Quando não havia nenhum mecanismo de prevenir contatos com nós egoístas, a taxa de entrega de mensagens sofria degradações. Em (Toh et al., 2010), um resultado similar foi alcançado, demonstrando que a taxa perda de mensagens aumentava até 500% quando a quantidade de nós egoístas na rede varia de 0% a 40% escolhidos aleatoriamente. Suponha, por exemplo, que o esquema de roteamento seja o repasse *two-hop*, no qual a mensagem é replicada em no máximo dois saltos a frente através de nós intermediários, assim, se uma mensagem é transmitida para um nó egoísta, a mensagem não é retransmitida, sendo então perdida. Portanto, o egoísmo é um fator que degrada severamente o repasse em redes oportunistas.

Portanto, a detecção de nós egoístas precisa e rápida é essencial para evitar a perda do desempenho geral da rede oportunista. Um grande desafio nesse ponto é a realização da detecção devido, principalmente pela perda frequente de conexão entre os nós e, a ausência de uma autoridade controladora, torna muito complexa a tarefa de detecção. De modo geral, as técnicas de detecção de comportamento egoísta se divide em dois tipos: baseados em crédito e baseados em modelos de reputação.

Enquanto abordagens baseadas em crédito exigem que os nós possuam algum tipo de crédito para participar de operações na rede, modelos baseados em reputação

atribuem valores para os nós da rede, no qual nós que cooperam mais possuem um maior ranqueamento, enquanto nós que agem de modo egoísta têm valores menores no ranqueamento.

No estudo de métodos para detecção de egoísmo em redes oportunistas, eventuais erros na detecção de nós egoístas são motivadores para criação de modelos que possam lidar com algumas detecções geradas erroneamente. Para fins de conhecimento, um erro de detecção ocorre quando um nó é sinalizado como egoísta e não é egoísta, e vice-versa. Estes fatores também podem degradar o desempenho geral da rede quando mecanismos de controle de comportamento não são bem empregados, como por exemplo detectar um nó como egoísta erroneamente e repassar essa informação adiante na rede sem realização de algum filtro. Neste ponto, ressaltamos que falhas são oriundas da arquitetura geral da rede oportunista como intermitência e mobilidade.

Este trabalho propõe um mecanismo de detecção de nós egoístas utilizando modelos de reputação que são capazes de prover uma avaliação confiável de um nó egoísta, conforme o número de avaliações aumenta sobre o nó. Para realizar a classificação, utilizamos uma técnica de aprendizagem de máquina conhecida como agrupamento (Jain & Dubes, 1988). Diferente de outras abordagens na literatura, esta proposta com classificação utilizando agrupamento é inovadora pois em trabalhos anteriores o modelo de reputação era aplicado com a utilização de limiares, isto é, quando um valor de ranqueamento ultrapassa um valor de limiar estipulado, este é classificado. Assim, nosso método difere pelo fato de, mesmo com poucas informações adquiridas, nosso modelo é capaz de classificar um nó, dado seu valor de ranqueamento, no grupo de nós egoístas ou não. Desse modo, também garantimos a classificação de todos os nós conhecidos por um vizinho em um agrupamento.

Para produzir o modelo de reputação introduzimos um modelo estatístico que emula uma competição onde, os nós recebem valores de reputação baseados na premissa de que durante um contato, um nó pode ser detectado como egoísta ou não egoísta com uma probabilidade de eficiência. Além disso, nosso método leva em conta o atual estado de conhecimento dos nós da rede sobre seus vizinhos. Assim, a atualização da reputação é dada em relação à reputação já conhecida de outros nós. Desse modo, tentamos criar um mecanismo que seja independente do número de nós egoístas na rede. Pois, quanto mais nós egoístas na rede, a cooperação gera uma maior pontuação, motivando o comportamento cooperativo.

1.1 Motivação

O contato é um evento singular em redes oportunistas. A oportunidade de um contato, que ocorre de forma diversa dependendo da topologia da rede é a principal premissa para o sucesso do roteamento de mensagens entre os membros da rede. Considerando que, para atingirmos o desempenho máximo de redes oportunistas são necessárias duas premissas: recursos ilimitados e colaboração dos nós da rede, o desempenho é sempre degradado quando uma ou ambas as premissas não estão presentes.

Embora muitos algoritmos de roteamento e encaminhamento projetados para redes oportunistas lidem com a utilização de recursos no projeto de desenvolvimento, normalmente eles não são projetados levando-se em consideração a falta de desejo dos nós de repassar mensagens, durante o contato com outros nós.

Por outro lado, o comportamento da não colaboração pode ser evitado ou mitigado, quando durante um contato, um dos nós identifica ou tem prévio conhecimento que o nó vizinho é egoísta. Neste caso, o conhecimento prévio pode auxiliar seu próprio mecanismo de repasse, além de poder informar previamente o restante da rede sobre o estado de comportamento dos nós. Assim, a tarefa de detecção também é acrescida da responsabilidade de sinalização e difusão das informações obtidas.

Dentro desse cenário misto de colaboração e não colaboração, modelos de reputação e de predição de contatos egoístas são itens fundamentais para o funcionamento eficaz de projetos de algoritmos voltados para DTN.

1.2 Objetivos

Este trabalho visa propor e validar um mecanismo baseado em modelos de reputação para predição de nós egoístas em redes oportunistas. Para atingir esse objetivo foi utilizado um modelo estatístico baseado em competição de usuários onde em nosso sistema há dois métodos de avaliação: cooperação ou não cooperação. Assim, esse modelo nos disponibiliza uma abordagem de ranqueamento, pelo qual os nós com históricos de colaboração mais frequentes tendem a obter valores maiores de pontuação, ao contrário de nós que apresentam histórico de comportamento egoísta.

Diferente de outros trabalhos na literatura (Marti et al., 2000; He et al., 2004; Bigwood & Henderson, 2011; Hernández-Orallo et al., 2015) que utilizaram limiares para definir um nó como egoísta ou não em modelos de reputação, nossa classificação dos nós em egoístas a partir do valor de reputação foi fundamentada em classificadores que utilizam agrupamentos, técnica conhecida também como clusterização. A razão

disso é acreditarmos que podemos antecipar a classificação de nós egoístas sem que estes tenham reputações que alcancem limiares predefinidos.

O objetivo principal é demonstrar que nosso modelo de reputação é capaz de distribuir bem as reputações dos nós mesmo quando há uma taxa de erro associada à detecção em cada contato. Além disso, nossa proposta tem o objetivo de classificar nós egoístas sem que haja uma grande quantidade de classificações erradas.

1.3 Contribuições

As principais contribuições deste trabalho são as seguintes:

- Implementar um modelo de reputação utilizando princípios estatísticos que sejam capazes de realizar uma boa separação entre reputações de nós egoístas e não egoístas, em tempo hábil para detecção e resiliente a quantidade média de nós egoístas na rede.
- Realizar a classificação utilizando um método de agrupamento e não através do uso de limiares. Desse modo, a classificação necessita de menos informação para predizer corretamente o comportamento dos nós da rede.
- Avaliar a eficácia da classificação via agrupamento usando o método do coeficiente de silhueta.
- Viabilizar prospectos para estudo futuro na área de detecção de nós egoístas em redes oportunistas.

1.4 Organização do Trabalho

Este trabalho está organizado da seguinte maneira: no Capítulo 2 apresentamos os principais conceitos relacionados a Redes Tolerantes a Atrasos e Desconexões, além dos trabalhos relacionados e conceitos sobre mobilidade, egoísmo e detecção de egoísmo em redes oportunistas. No Capítulo 3 é apresentada a arquitetura de rede utilizada neste trabalho bem como as definições usadas para o projeto de detecção, construção, atualização do modelo de reputação e como é realizada a classificação do nós. No Capítulo 4 é apresentado o detalhamento da implementação realizada, os cenários utilizados para simulação, as métricas de avaliação e, por fim, a discussão dos resultados. Finalmente, nos Capítulos 5 e 6 são apresentadas as conclusões tiradas e possíveis trabalhos futuros neste campo de estudo.

Capítulo 2

Conceitos e Trabalhos Relacionados

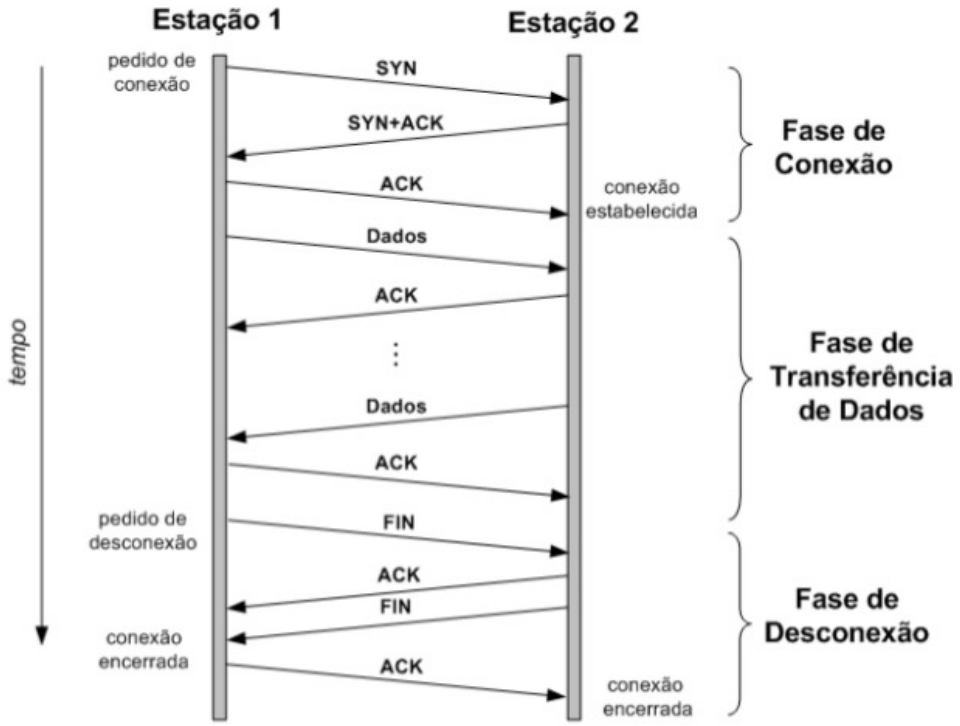
Neste capítulo apresentamos e descrevemos os principais conceitos, arquitetura, tipos de aplicações existentes e os principais problemas relacionados às redes tolerantes a atrasos e desconexões. Além disso apresentamos um compacto do estado da arte sobre egoísmo em redes DTN e redes oportunistas.

2.1 Redes Tolerantes a Atrasos e Desconexões

O esquema clássico de funcionamento da *Internet* está baseado no modo operacional do protocolo de transporte TCP (*Transmission Control Protocol*), um protocolo responsável por garantir confiabilidade na comunicação fim a fim. Com o TCP, as premissas de comunicação confiável são mantidas independentes da infraestrutura usada nas subredes, não havendo preocupação com o tipo de tecnologia de canal de comunicação usada (ex: fibra, coaxial, par trançado, radiofrequência, etc.) sob o qual o protocolo IP (*Internet Protocol*) está operando (Oliveira et al., 2007). Este esquema consiste de uma sequência de passos: o estabelecimento da conexão, a transferência de mensagens e a desconexão como ilustrados na Figura 2.1. Toda sequência de informação transferida entre duas estações (de transmissor para receptor) na rede é sinalizada como boa, isto é, quando o envio ocorre completamente e sem falhas, sempre que há um reconhecimento positivo do receptor na comunicação, um ACK (*acknowledgment*), que indica que a comunicação ocorreu com sucesso.

Contudo, determinados ambientes possuem características que tornam árdua a tarefa de encaminhamento de mensagens sob a tutela do modelo TCP/IP tradicional devido a alguns fatores que impossibilitam a conexão contínua entre dispositivos envolvidos, tais como distância, mobilidade, energia limitada, defeitos, entre outros. Exemplos de ambientes onde a comunicação não pode ser realizada de forma contínua

Figura 2.1: As fases da operação TCP.



Fonte: (Oliveira et al., 2007).

e com baixos atrasos incluem: redes de comunicação militares, espaciais, redes *ad hoc* móveis, etc. Nestes cenários as trocas de mensagens sofrem com a variação da largura de banda, desconexão dos nós *hosts*, atrasos e mobilidade, causando perda de desempenho no modo de funcionamento da pilha de protocolos TCP/IP. É nesse contexto que emergiram as redes DTN.

Ao passo que nas redes infraestruturadas os nós da rede podem se comunicar por um caminho fim a fim, não há necessidade de armazenar as mensagens através dos dispositivos intermediários da comunicação. Contudo, devido as iminentes desconexões, as redes DTN utilizam uma técnica conhecida por armazenamento persistente. Quando uma mensagem precisa ser transmitida para um nó da rede, este é copiado e repassado para o destinatário através dos contatos oportunistas dos nós da rede, que armazenam as mensagens e repassam para nós que não estejam carregando aquele dado. Esta técnica é chamada de armazena e encaminha (*store and forward*), ou seja, uma mensagem é transmitida através de um nó intermediário, que recebe e mantém a mensagem em seu *buffer* e, retransmite-a em um próximo contato oportunista para o destinatário final ou para outro nó intermediário. Tal como ilustrado na Figura 2.2,

o nó S envia mensagens através de comunicação oportunista, apresentada pelas setas não tracejadas, para os nós intermediários até a mensagem alcançar o destinatário final D.

Assim, podemos resumir a arquitetura DTN a partir de três mecanismos:

- **Transferência em custódia**: se uma mensagem M precisa ser transmitida de uma origem S até o destino D, cujo caminho fim a fim é inexistente, tal mensagem é custodiada, isto é, entregue como cópia, a nós intermediários que S encontra de modo oportunista.

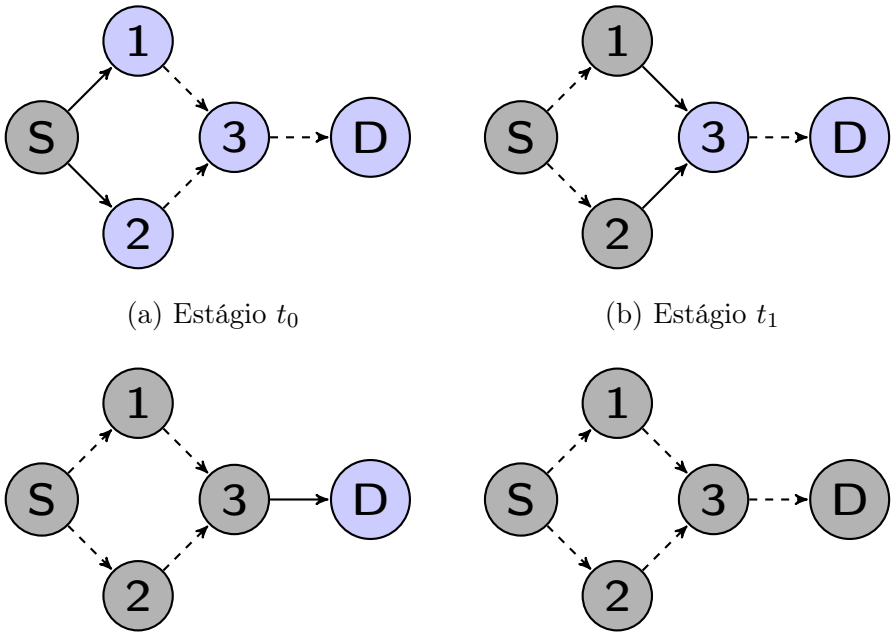
- **Roteamento de mensagens específicas**: devido ao baixo conhecimento sobre caminhos possíveis em redes DTN, o roteamento é feito de modo específico, assumindo-se estratégias em custódias fixas ou ainda analisando o comportamento da rede para prever melhores transferências em custódias futuras que viabilizem a entrega das mensagens ao destinatário final.

- **Periodicidade**: devido à mobilidade aleatória dos nós da rede, é necessário pressupor alguma periodicidade no comportamento para que não haja infinitas configurações no mesmo sistema, dificultando o roteamento e a transferência em custódia. Devido aos padrões de mobilidade social já identificados na literatura é possível afirmar que há alguma periodicidade nos ambientes DTN existentes.

Deve ser ressaltado que a operação *store and forward* em DTN difere de outros protocolos tradicionais de rede. Por exemplo, em redes IP baseadas nesta operação, o armazenamento ocorre por um pequeno período de tempo até que as mensagens sejam encaminhadas para o próximo nó. Esse armazenamento é normalmente feito por memórias dinâmicas (ex: chips de memória de roteadores) e o tempo de armazenamento é da ordem de milisegundos. Em contraste, como as DTN não operam sobre enlaces que estão sempre disponíveis, é esperado que os nós armazenem mensagens durante algum tempo. Nesse caso, o tempo de armazenamento em DTN pode ser até da ordem de horas ou dias, sendo preciso alguma forma de armazenamento persistente e robusto (ex. disco rígido, memória flash de dispositivos portáteis) para preservar as informações diante de reinicializações no sistema (Oliveira et al., 2007).

Como DTN opera sob o esquema de *store and forward* então é necessária uma adaptação na implementação da pilha de protocolos tradicionais TCP/IP, de modo que este suporte as premissas de DTN. Assim, para realização da comunicação oportunista, devem ser implementados métodos que estabeleçam o armazenamento persistente e a transmissão de mensagens de modo oportunista nas próprias aplicações, tornando o

Figura 2.2: Visualização da estratégia de *Store and Forward* para o encaminhamento de mensagens entre os nós S e D.



(a) Estágio t_0 (b) Estágio t_1

(c) Estágio t_2 (d) Estágio t_3

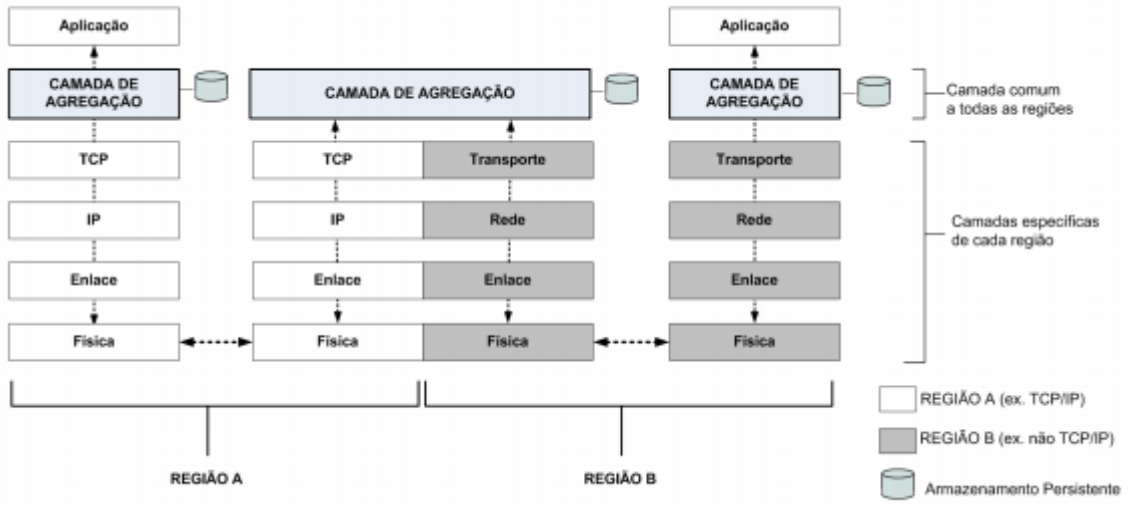
Fonte: o autor.

funcionamento de cada dispositivo similar ao de *gateways*, com tabelas de repasse e gerência do armazenamento. No entanto, isso tornaria o processo de adaptação de aplicações despendioso, visto que cada aplicação teria que ser desenvolvida levando em consideração problemas de frequentes desconexões ou atrasos. Para permitir que a comunicação seja feita de forma transparente e independente das estruturas de aplicações a solução encontrada foi criar uma camada adjacente as camadas de transporte e aplicação, a camada de agregação (*bundle layer*), responsável por encapsular as mensagens (*bundles*) e manter a condição de interoperabilidade das redes baseadas no modelo OSI tradicional e cujo as especificações estão documentadas na RFC 5050 (Scott, K., Burleigh, 2007).

O fluxo de comunicação entre camadas abaixo da camada de agregação são ilustrados na Figura 2.3 e são definidas de acordo com a conveniência do ambiente de comunicação de cada região (canal cabeado, canal sem fio, canal *ad hoc*), podendo ser específicas para cada região englobada pela DTN. Esta configuração permite que a arquitetura DTN comunique-se de forma transparente diante da configuração heterogênea dos variados tipos de rede, caracterizada por múltiplas interfaces de comunicação.

O protocolo de agregação tem como função gerar as mensagens que servirão de

Figura 2.3: A camada de agregação da arquitetura DTN.



Fonte: (Oliveira et al., 2007).

controle entre as aplicações DTN. Estas mensagens são transformadas, na camada de agregação, em uma ou mais unidades de protocolo (*Protocol Data Units* – PDUs) denominadas agregados (*Bundles*), que armazenam dados de aplicação de usuário e provêem as informações necessários do nó remetente para o nó destinatário das mensagens. Os *bundles* são encaminhados pela rede através dos nós DTN em um processo denominado encaminhamento (*forwarding*). Assim, nestes ambientes, por exemplo, o pedido de transferência de um arquivo pode ser transmitido contendo informações do usuários remetente, o nome de um arquivo de interesse, etc. Todas essas informações são compactadas em um *bundle* e transmitidas de uma única vez de forma persistente, evitando a sequência de interação presente em uma rede TCP/IP tradicional. Os *bundles* podem ser eventualmente fragmentados para facilitar o encaminhamento na rede.

O repasse de mensagens em uma DTN pode ter as seguintes opções (Neto, 2011):

- **Número de cópias**: estabelece quantas cópias devem ser transmitidas pelo remetente.
- **Transferência de custódia**: transfere a responsabilidade de armazenamento da mensagem.
- **Recibo de transferência de custódia**: o nó encaminhador informa ao remetente a aceitação da custódia.

- **Recibo de entrega fim a fim**: o destinatário informa ao remetente o recebimento da mensagem.
- **Recibo de exclusão**: é informado ao remetente a exclusão da mensagem.
- **Recibo de encaminhamento**: o nó encaminhador informa ao remetente o encaminhamento da mensagem.
- **Recibo de recepção**: o nó encaminhador informa ao remetente o recebimento da mensagem.

O repasse, no entanto, tem uma estrita relação com a mobilidade dos nós da rede. Esta relação leva o modo como os nós interagem entre si ao longo do tempo. Foi nesse âmbito que surgiram as redes oportunistas, uma variação das rede DTN, no qual os dispositivos são carregados por usuários humanos e no qual a comunicação ocorre através do repasse pelos contatos oportunistas que ocorrem durante o tempo.

Nas subseções seguintes são apresentados alguns conceitos relevantes sobre redes oportunistas que nos ajudam a compreender o problema tratado neste trabalho como a mobilidade, padrões sociais e contato entre os nós em redes oportunistas e, por fim, o egoísmo em redes oportunistas.

2.1.1 Mobilidade em Redes Oportunistas

Como discutido na seção anterior, o repasse de mensagens em DTN depende de como os nós se comunicam. O contato oportunista é o item base para criação de modelos otimizados de encaminhamento e está ligado diretamente a mobilidade dos dispositivos. De um modo geral, modelos de mobilidade podem ser divididos em duas categorias (Liu et al., 2011): modelos sintéticos e modelos extraídos do mundo real (Mouly & Pautet, 1992). Modelos sintéticos estão relacionados à mobilidade gerada através de formulações matemáticas ou eventos estocásticos, enquanto que modelos baseados em *traces* (descrição de contatos entre os nós da rede) são relatados a partir de dados de mobilidade extraídos em cenários reais e podem representar objetos como movimentação humana ou de carros, por exemplo.

2.1.1.1 Modelos de mobilidade Sintéticos

Os modelos de mobilidade sintéticos representam formulações matemáticas para apresentar modos de movimentação em uma determinada superfície. Estes modelos são utilizados para avaliação de desempenho e comparação de diversos algoritmos em DTN,

como algoritmos de roteamento ou controle de *buffer*. Abaixo apresentamos dois modelos particulares de mobilidade consagrados na literatura.

Modelo *Random Waypoint*

O modelo de mobilidade *Random Waypoint* foi apresentado por Johnson e Maltz em (David B. Johnson, 1996). Nesse modelo, cada nó é colocado inicialmente em uma posição aleatória (*waypoint*) dentro de uma área de interesse $N \times M$, e então movimentam se ao longo desta superfície para o próximo *waypoint* com velocidade aleatória entre $[Speed_{min}, Speed_{max}]$, escolhida a partir da distribuição de probabilidade uniforme. Quando este alcança o próximo *waypoint*, permanece neste ponto durante um período de tempo chamado de tempo de pausa. Quando o tempo de pausa expira, repete-se o processo de movimentação acima.

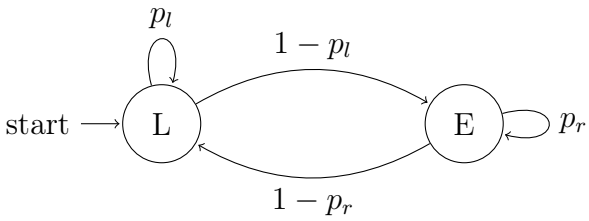
Este modelo foi amplamente utilizado na literatura para diversos comparativos em DTN (por exemplo em (Abdulla & Simon, 2007; Wang et al., 2013; Luo et al., 2008; Krifa, 2012)). Contudo, esse modelo, quando utilzado em avaliação de desempenho, pode não representar com fidelidade uma gama de peculiaridades da mobilidade humana, por exemplo. A seguir é apresentado um modelo sintético que tenta representar estas ditas peculiaridades do movimento humano, o modelo de mobilidade baseado em comunidade.

Modelo baseado em Comunidades

Diversos estudos (Furletti et al., 2013; Le et al., 2014; Crowcroft, 2008; Cho et al., 2011; Boldrini & Passarella, 2010; Karagiannis et al., 2010) analisaram características da mobilidade humana baseado em *traces* reais e avaliaram que humanos tem a tendência de visitar determinados locais mais frequentemente que outros como: universidade, trabalho, casa, etc. Kim et al. em (Kim et al., 2006) conduziram um experimento de aproximadamente 12 meses no campus de Dartmouth. Essa investigação mostrou que os estudantes passavam a maior parte do tempo em localizações específicas do campus como cafeteria ou biblioteca. Foi demonstrado também que os movimentos humanos seguem padrões de relacionamento humano. Estas observações corroboram as pesquisas já relatadas acima.

Baseado nestas observações, o modelo baseado em comunidade foi apresentado em (Spyropoulos et al., 2006). Neste modelo, cada nó é associado a uma célula geográfica, chamada de região local, considerada preferida desse nó. A depender do movimento interno do nó, a mobilidade deste pode ser definida como um conjunto de épocas locais e de saída. A movimentação local trata do movimento aleatório dentro da célula, enquanto que a saída significa o período quando o nó saiu de sua célula para andar no

Figura 2.4: Máquina de estados do modelo de mobilidade baseado em Comunidade.



Fonte: o autor.

restante da área. Se a posição anterior do nó era local, então a próxima posição será uma posição local com probabilidade p_l ou uma posição fora da célula com probabilidade de $1 - p_l$. Similarmente, se tivermos a situação oposta, a próxima posição será a posição externa com probabilidade p_r e local com $1 - p_r$ como apresentado pela máquina de estados ilustrada na Figura 2.4.

2.1.1.2 Traces de Mobilidade

Traces são uma coletânea de dados sobre contato ou sobre localização extraídos a partir de experimentos em ambientes reais. Estes *traces* podem ser provenientes de duas variantes: *traces* de localização e *traces* de contato.

Os *traces* de localização são extraídos a partir de medições geográficas de sistemas de GPS (*Global Positioning System*). São um conjunto de dados que contém informações das posições geográficas dos participantes desses experimentos (McNett & Voelker, 2005; Lee et al., 2008, 2009; Azevedo et al., 2009; CRAWDAD, 2014; USC, 2014).

Ao passo que *traces* extraídos de GPS podem representar *datasets* mais esparsos, *traces* de contato representam medidas de proximidade entre os objetos da rede, além de ser uma ferramenta de medida mais eficiente para avaliação de similaridade entre os objetos da rede. Alguns *traces* foram obtidos através do escaneamento de contatos próximos utilizando tecnologia *bluetooth* ou *Wi-fi Direct* em diferentes ambientes como conferências, locais urbanos, campus universitários, etc. Na Tabela 2.1 é apresentada algumas informações relevantes sobre os traces pesquisados durante esse trabalho: MIT *Reality Mining* (Eagle & (Sandy) Pentland, 2005), Rollernet (Tournoux et al., 2009; Rollernet, 2014), Cambridge, INFOCOM 05 e INFOCOM 06 (Haggle, 2014).

Traces de mobilidade têm sido largamente estudados com o intuito de propor algoritmos para redes oportunistas com base em informações menos voláteis, isto é, menos aleatórias como nos modelos de mobilidade sintéticos, e que representem a mo-

Tabela 2.1: Traces de contatos e suas características.

	Duração (dias)	# de dispositivos	Descrição
MIT	365	97	Campus universitário
Cambridge	11	54	Campus universitário
INFOCOM 05	3	41	Conferência
INFOCOM 06	3	98	Conferência
Rollernet	0.125	62	Evento social

bilidade humana como ela é, através do padrão de interação social entre participantes dos experimentos (Ii et al., 2013; Socievole et al., 2013; Wu et al., 2013; Zhu et al., 2014; Hui & Crowcroft, 2007; Hui et al., 2011). Do mesmo modo que o padrão de comportamento social dos nós na rede influencia na criação de novos métodos, há também outros fatores do comportamento social que têm despertado o interesse de pesquisa e que precisam ser considerados no projeto de novas metodologias para redes oportunistas. Um fator em particular que pode mudar o desempenho geral do projeto de uma DTN é o egoísmo presente nos nós da rede. Uma gama de pesquisas têm sido relatadas nesse campo e são descritas com maiores detalhes na seção 2.2.

2.2 Egoísmo em Redes Oportunistas

Além de considerarmos a mobilidade como descrito na seção anterior é necessário pensar nas autoridades responsáveis por viabilizar os objetivos principais da rede. De um modo geral, podemos definir os principais exemplos de DTN e suas autoridades conforme descrito na Tabela 2.2. Nas redes que possuem alguma autoridade, o incentivo é definido pelo grupo. Assim os nós são desenvolvidos (aqui ressalta-se que os nós nesses exemplos normalmente são máquinas que não sofrem interferência humana após a implantação) para executarem uma atividade comum definida pela autoridade. Entretanto nas redes que chamamos de Civis, não há uma autoridade responsável pela definição dos objetivos de rede. Entre esses tipos de rede destacam-se as redes formadas por dispositivos que podem sofrer influência humana como redes oportunistas de celulares.

No mecanismo comum da comunicação, nós servem como encaminhadores para outros nós ao carregar mensagens de forma intermediária. Obviamente, isto requer uma comunicação em modo totalmente cooperativo. Contudo, em aplicações reais, os nós (*smartphones*, por exemplo) em redes oportunistas podem ser objetos controlados por pessoas reais. Além disso, o processo de transmissão demanda energia ou espaço

Tabela 2.2: Exemplos de DTN e os tipos de cooperação.

Rede	Autoridade	Incentivo	Tipo de Aplicação
Militar	X	grupo	Comunicação no campo de guerra
Sensoriamento	X	grupo	Monitoramento e coleta de dados
Civil (rede oportunista)	–	individual	Comercial, entretenimento, veicular e doméstico

de *buffer*, que é limitado aos nós, então estas pessoas podem agir de forma egoísta e não desejando cooperar com o repasse de mensagem para os outros (Wu et al., 2012).

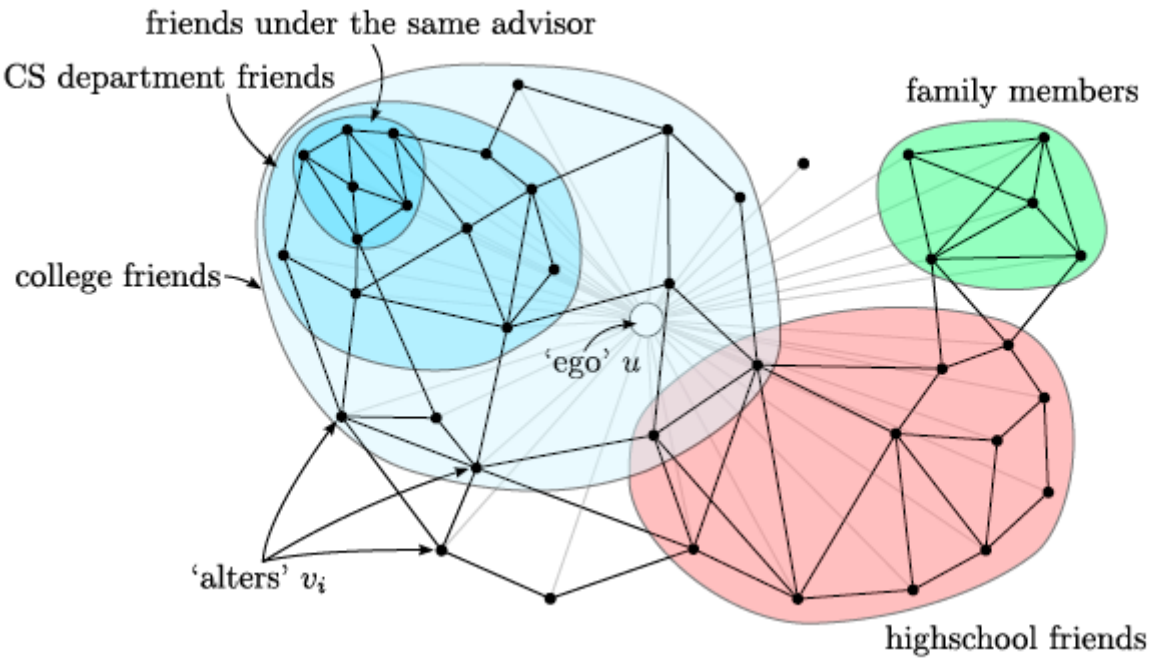
Neste ponto definimos o egoísmo em DTN como a não cooperação no repasse de dados na rede com o intuito normalmente de economizar recursos limitados dos seus dispositivos, assim os nós egoístas não têm como objetivo prejudicar outros nós da rede, criar informações, burlar o controle de fluxo e nem alterações das informações de roteamento.

O egoísmo em DTN ocorre por diversos motivos, principalmente como forma de economizar os recursos, limitados, dos nós da rede tais como consumo de memória primária/secundária, energia, banda, etc. Para compreender esse efeito, podemos nos voltar a sociologia e observar os mecanismos por quais as pessoas decidem com quem interagir. Empiricamente é observado o fato que o *status* social da vizinhança de um nó é um bom indicador de seu próprio *status*. A razão é que interações sociais requerem recursos que são limitados, então os seres humanos escolhem cuidadosamente com quem compartilhar seus recursos, maximizando seus próprios recursos individuais (Pujol et al., 2009). De um modo geral, temos que as pessoas tendem a interagir mais com pessoas que compartilham mais gostos sociais que outras.

Uma forma de compreenda tal comportamento também é apresentado em (Souza et al., 2017), no qual um questionário foi aplicado aos alunos da Universidade Federal do Amazonas. Uma das perguntas foi: Se uma pessoa envia uma solicitação de conexão, porém a bateria do seu dispositivo está em um nível crítico, o que você faria? (a) aceitaria o convite (b) não aceitaria o convite (c) esperaria recarregar a bateria para depois comunicar com a pessoa. 10% das pessoas que responderam ao questionário aceitariam o convite, enquanto 55% das respostas indicaram que esperaria até o recurso de bateria ser recarregado.

Uma outra forma de visualizar a interação social humana é imaginar um grafo de relações sociais de determinada pessoa. Esta pessoa está socialmente ligada a um ou mais subgrafos sociais com os quais compartilha laços comuns com diferentes níveis de força de interesse. Na Figura 2.5, podemos ver como um nó u se comporta com outros amigos v_i e a classificação destes em seus respectivos *clusters* sociais, isto é,

Figura 2.5: Círculo social do usuário, 'ego', com outros amigos denominados de 'alters' e sua identificação social com o usuário.



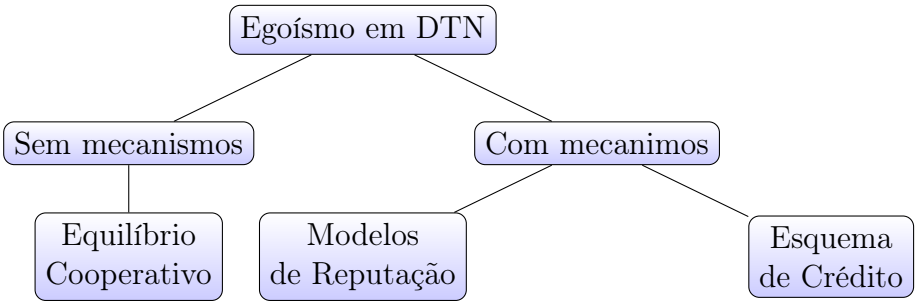
Fonte: (McAuley & Leskovec, 2012).

comunidades sociais.

Sob a condição de relacionamentos podemos dividir o egoísmo de duas maneiras. Se um nó demanda por economizar algum determinado recurso e ele não coopera com a comunicação nem mesmo dentro do mesmo grupo social, então temos o egoísmo individual (*individual selfish*), quando um nó demanda recursos compartilhados da rede, mas negando compartilhar seus próprios recursos com os demais. De outro modo, quando um nó coopera mais com usuários de um mesmo grupo social que ele, mas não auxilia no repasse com outros de fora do seu grupo social, então temos o egoísmo social (*social selfish*). O egoísmo social não representa uma contradição ao egoísmo individual, mas uma extensão em si, visto que, quando um nó não tem laços sociais, seu egoísmo social torna-se egoísmo individual.

Aliado a essa condição, alguns estudos na literatura (Wu et al., 2012; Karaliopoulos, 2009; Li et al., 2012; Panagakis et al., 2007; Li et al., 2011) avaliaram o impacto do comportamento egoísta sob diferentes métricas de desempenho em protocolos de DTN como taxa de entrega ou atraso médio e demonstraram que sob a ótica egoísta dos usuários da rede, o desempenho destes protocolos são severamente degradados, quando uma porção dos nós é egoísta ou quando estes têm probabilidade alta de não

Figura 2.6: Tipos de tratamento de nós egoístas.



Fonte: o autor.

cooperar com a comunicação. Isso ocorre principalmente porque muitos protocolos, principalmente protocolos de roteamento partem do pressuposto que todo nó pode e irá colaborar com a transmissão durante o contato. Estes resultados demonstraram a necessidade de métodos que possam facilitar a transferência e gerência de dados em redes sob estas condições de comportamento dos nós da rede.

2.2.1 Detecção de Comportamento Egoísta

Com relação ao egoísmo, alguns estudos analisam de formas distintas a maneira como o egoísmo deve ser lidado. Buttyan *et al.* argumentam a possibilidade da conectividade da rede ser mantida sem a necessidade de mecanismos externos (Buttyan et al., 2005). A razão seria a dependência mútua existente entre os nós da rede que induziria a um equilíbrio cooperativo sobre o comportamento dos nós. Dessa forma, pelo menos uma estratégia cooperativa será seguida pelos nós, assim não necessitanto de um mecanismo externo para detecção.

Por outro lado diversos métodos têm sido propostos para detectar comportamentos egoístas em DTN e MANET (*Mobile Ad Hoc Networks*) através de mecanismos externos. Estes esquemas podem ser divididos em duas categorias: baseado em crédito e baseado em reputação. Para sumarizar, a abordagem baseada em crédito visa desencorajar um nó de ser egoísta a partir de um sistema de créditos, no qual apenas nós com crédito conseguem participar ou participar mais da rede de dados. O sistema de detecção baseado em reputação visa detectar o nó egoísta e ir admitindo valores de reputação, que refletem o comportamento deste durante o repasse, assim o comportamento egoísta é avaliado conforme o histórico de comportamento de cada nó na rede. Um breve diagrama destes métodos é apresentado na Figura 2.6.

Adicionalmente, nos modelos baseados em reputação, o auxilio de um método de detecção se faz necessário, visto que estes são modelos compostos, no qual a composição

é feita pelo sistema de monitoramento (também chamado de detecção) e o modelo de reputação, responsável por agregar um valor histórico para as observações oriundas do sistema de detecção.

Mais detalhes sobre o funcionamento de esquemas de créditos são exibidos na subseção 2.2.1.1. Uma amostra de estudos na literatura sobre métodos de detecção e modelos de reputação é abordada na subseção 2.2.1.2.

2.2.1.1 Esquema de Crédito

Na abordagem baseada em esquema de crédito, um mecanismo de incentivo ou de crédito monetário são aplicados. Os dispositivos da rede devem pagar para que suas mensagens sejam repassadas usando moedas virtuais ou crédito (Buttyán & Hubaux, 2003; Zhong et al., 2003; Miranda & Rodrigues, 2003; Crowcroft et al., 2004). Assim, o crédito tem como finalidade encorajar o dispositivo a participar das atividades de repasse na rede. Se um nó não repassa mensagens de outros nós da rede, ele não poderá coletar créditos suficientes para que as mensagens criadas por ele sejam repassadas por outros nós da rede. Dessa forma, o egoísmo é desencorajado como forma de obtenção de créditos necessários para o estabelecimento de repasses.

Buttyán *et al.* (Buttyán & Hubaux, 2003) propuseram o uso de uma moeada virtual, chamada *nuglets*. Quando um nó repassa uma mensagem, o contador de *nuglet* é incrementado por um. Quando esse nó requisitar que outro nó repasse sua mensagem, ele deve possuir uma certa quantidade de *nuglets* que seja maior que a quantidade de nós intermediários necessários para que a mensagem seja repassada para o destinatário final. Assim, se o nó que requisita o repasse de sua mensagem não possuir crédito suficiente, sua mensagem não será repassada através da rede. Desse modo, o nó precisa atuar como nó intermediário de outros nós para conseguir crédito suficiente para que suas mensagens sejam repassadas na rede.

Zhong *et al.* (Zhong et al., 2003) propuseram o mecanismo *Sprite*, um sistema de crédito à prova de fraude para uso de modo centralizado. Neste mecanismo, cada nó mantém uma gravação sobre recepção de mensagens, isto é, para toda mensagem recebida por esse nó, ele guarda informações sobre nós envolvidos no repasse desta mensagem, os nós intermediários pelo qual a mensagem passou. Assim, a cada certa janela de tempo, o nó reporta estes dados para um serviço de gerência, que determina a quantidade de cobranças ou crédito para todos os nós envolvidos nas transações de mensagens. A principal desvantagem deste mecanismo é o uso de uma autoridade central, que é um ponto de falha, principalmente em redes compostas por dispositivos carregados por pessoas, no qual a atribuição de autoridades é uma tarefa complexa.

Miranda *et al.* (Miranda & Rodrigues, 2003) propuseram um algoritmo que mantém uma vigilância sobre o estado dos nós vizinhos. Cada vizinho é classificado de acordo com três estados: amigo, inimigo ou egoísta. Cada nó monitora continuamente o estado de seus vizinhos e troca periodicamente informações com seus vizinhos contendo informações sobre o estado de seus vizinhos. Além disso, é proposto um algoritmo descentralizado para evitar esquemas complicados de pagamento de crédito, entretanto, o overhead apresentado neste esquema é alto devido as periódicas trocas de mensagens por todos os nós da rede.

Em (Crowcroft et al., 2004), Crowcroft *et al.* propuseram um mecanismo de crédito que leva em consideração fatores como banda de rede e energia para modelar o esquema de incentivo. A energia e banda utilizadas são adaptadas usando um modelo de otimização para controle de taxas. Assim, cada nó atualiza o preço pelo uso de seus recursos baseados na energia atual do nó e uso de banda.

Srinivasan (Srinivasan et al., 2003) propôs usar uma estratégia baseada no conceito de *Generous Tit-For-That* (GTFT) para realizar um processo de decisão binário, aceitar ou rejeitar o repasse de uma mensagem vindo de outro nó vizinho com o objetivo de otimizar o *throughput* da rede.

Em síntese, estes algoritmos visam resolver o problema de cooperação em redes *ad hoc*, mas eles também introduzem um grau de complexidade para os nós da redes tais como *overhead* de pacotes de controle ou a implantação de módulos de segurança auxiliares para evitar fraudes na geração de créditos. Além disso, estes esquemas podem não ser justos com certos nós da rede como nós que possuem poucos contatos com nós da rede, por exemplo, desde que o número de contatos pode não ser suficiente para adquirir crédito suficiente para que suas mensagens sejam repassadas por outros nós. Estes problemas ainda não solucionados para estes tipos de mecanismos tornam o esquema baseado em crédito um esquema complexo de ser implantado em redes oportunistas reais.

2.2.1.2 Modelos de Reputação

Redes auto-organizadas tais como redes oportunistas, MANET, redes P2P (*Peer-to-peer*) e redes em malha têm como premissa a cooperação para tirar o proveito máximo do desempenho da rede. Contudo, usuários estão primariamente priorizando o seu próprio benefício e colocando o fator colaborativo em segundo plano. Portanto a relação entre a cooperação e o desempenho ótimo não podem ser garantidos. Esse é um fenômeno conhecido como *free-rider problem*[1], um conhecido problema econômico que

[1]https://plato.stanford.edu/entries/free-rider/

refere-se ao fato de alguém beneficiar-se de um recurso ou bem de consumo sem pagar uma taxa para utilização deste benefício.

Um modo interativo e distribuído de realizar o controle de usuários com a característica de *free-rider* é computar o valor de reputação sobre seu subgrafo de contatos oportunistas (MANET e DTN) ou sobre seus *peers* (P2P), por exemplo. Dessa forma, tais valores podem ser utilizados para decidir que nós estão cooperando com a comunicação e que nós estão agindo de modo egoísta. Em resumo, esses mecanismos executam as funções de monitoramento (também chamado de detecção), construção do modelo de reputação e tratativas de respostas aos comportamentos indesejados.

O uso de modelos de reputação já tem sido provado como uma solução popular e eficaz em sistemas online como *eBay*, onde para cada par de transação é atribuída um nota, também chamada de *rating*, em redes sociais como *Stack Overflow* para avaliar melhores respostas, em jogos onlines para medir o desempenho de jogadores[2] em competições de programação[3] e até mesmo é utilizado no *rank* FIFA™de seleções[4] para mensurar a força de seleções de futebol. Contudo, diferentemente das redes auto-organizadas estas aplicações possuem a propriedade de serem centralizadas, tornando seus métodos, da maneira como são projetados e executados, impossíveis de serem aplicados no tipo de rede do escopo deste trabalho que exige um método independente de uma autoridade central, no qual cada nó opera de maneira independente e sem o total conhecimento das operações da rede.

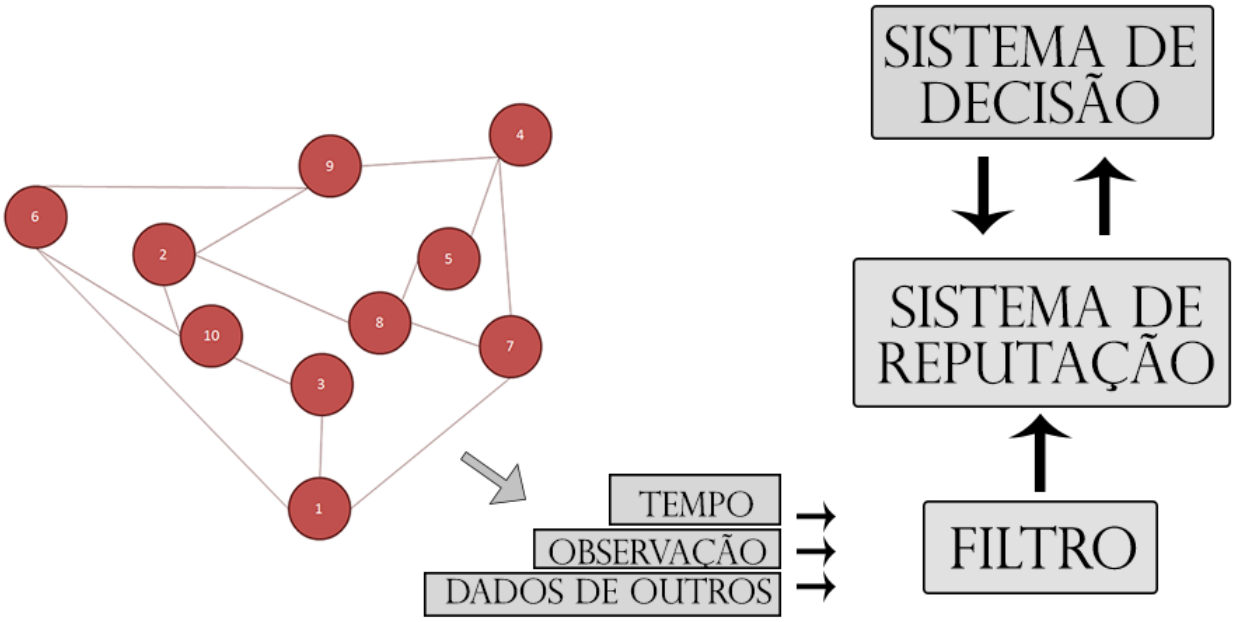
Assim, para realização de um sistema de predição de comportamento futuro em redes oportunistas é preciso que ele compute informações passadas. De modo não formal, um sistema de predição precisa manter informações passadas sobre um determinado conjunto de outros usuários para prover uma base sólida para um processo de decisão. Esse sistema deve ser capaz de receber parâmetros como o tempo em que foi observado um dado, valor da observação própria e observação de outros usuários como ilustrado na Figura 2.7, onde várias fontes contribuem para o modelo de reputação ideal: observações próprias, de outros usuários e o tempo de coleta.

Além disso, é preciso que ele funcione de modo distribuído para melhor desempenho da predição. A premissa básica por trás deste pensamento é conhecida como efeito *gossip*, no qual as informações propagam rapidamente quando outros agentes colaboram na propagação da informação. Embora o efeito *gossip* seja vulnerável a propagação de falsas informações, este trabalho não abordará esta linha de estudos.

Dado o fato que o campo de estudos sobre egoísmo abrange basicamente detecção,

[2]http://www.pcgamer.com/csgo-ranks-explained/
[3]http://codeforces.com/blog/entry/102
[4]http://www.eloratings.net/about

Figura 2.7: Fluxo de um modelo de reputação.



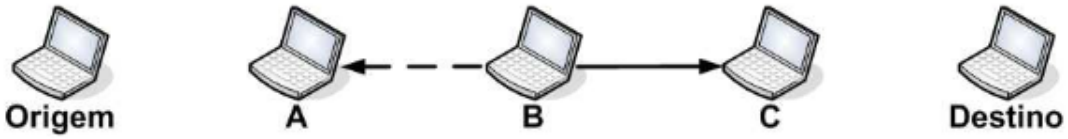
Fonte: o autor.

modelos de reputação e roteamento, nós focamos aqui apenas em detecção e modelos de reputação pois este trabalho aborda uma estratégia modelo de reputação em DTN independente do roteamento utilizado. Ao passo que a reputação é intrinsecamente ligada a detecção de nós egoístas também abordamos algumas estratégias de detecção e estratégias híbridas que incluem detecção e modelos de reputação apresentadas na literatura.

Um dos primeiros mecanismos para detecção de nós egoístas foi proposto por Marti *et al.* em (Marti et al., 2000). Eles propuseram um mecanismo de vigilância, o *watchdog* e um mecanismo de controle chamado *pathrater*. Seu funcionamento consiste da seguinte maneira: quando um nó A repassa uma mensagem para um nó B, um terceiro nó *watchdog* analisa a comunicação e verifica se durante um período de tempo o nó B irá repassar a mensagem ou não e se este não repassar, então é incrementado um índice de falha de cooperação para este nó. Quando este nó tem um índice de falha que ultrapassa uma determinada marca, então este determina que o nó apresenta mal comportamento e repassa a mensagem para os outros nós da rede.

O funcionamento do *watchdog* é detalhado na Figura 2.8. É assumido que existe um caminho entre os nós Origem e Destino, passando pelos nós intermediários A, B e C. O nó de origem então envia uma mensagem para o nó A, que ao recebê-lo repassa para o nó B. Após repassar a mensagem, o nó A inicia o monitoramento para verificar se B repassou a mensagem. Esse monitoramento tem como objetivo observar se o nó B encaminha corretamente a mensagem para o próximo salto da rede, ou seja, para o nó C. Caso a mensagem não seja encaminhado, o nó A classifica esse comportamento do nó B como egoísta.

Figura 2.8: Funcionamento do *Watchdog*.



Fonte: adaptado de (Marti et al., 2000).

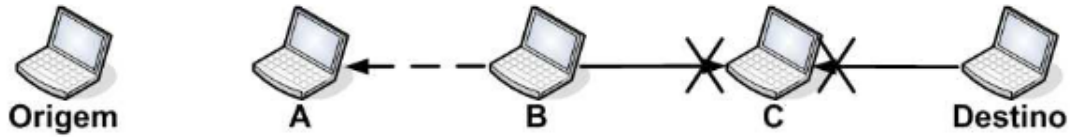
O *watchdog* utiliza um *buffer*, que armazena as mensagens por um período de tempo. Ao receber cada mensagem observada pelo modo promíscuo da camada de rede, o *watchdog* o compara com as mensagens contidas neste *buffer*. Caso alguma mensagem seja igual a outra contida no *buffer*, então essa é removido. Caso a mensagem permaneça no *buffer* até expirar um determinado período de tempo, um contador de eventos é acionado, significando que o outro nó não repassou determinada mensagem. Quando este contador atinge um limiar, o nó é detectado efetivamente como egoísta.

O mecanismo de controle da reputação, o *pathrater*, é também o gerenciador de rotas do mecanismo citado acima. O *pathrater* está ligado a todos os nós da rede e agrega o conhecimento dos nós egoístas ou maliciosos da rede com as rotas de confiança a fim de predizer a rota mais confiável. Isso é feito calculando a métrica do canal através da média das reputações dos nós desta rota. DTN, no entanto, possui uma quantidade fatorial de rotas, tal que cada rota nunca pode ser prevista com alta confiabilidade, tornando a estabilização deste método custoso para aplicação em DTN. Além disso, esse método apresenta outras desvantagens como *overhead* de mensagens trocadas entre os *watchdogs* e a complexidade no cálculo de limiares para o *pathrater*.

Outra dificuldade relacionada aos *watchdogs* está na geração de falsos positivos, isto é, detecções erradas. Um exemplo de uma detecção errada pode ser vista na Figura 2.9, no qual nó A percebe que nó B encaminhou a mensagem para C, no entanto C não recebeu a mensagem por motivos de *buffer* cheio, preferindo rejeitar a mensagem. Desse modo, podemos dizer que o *watchdogs* podem não assegurar o repasse de mensagens

ao destinatário. Isto ocorre ao fato de que na arquitetura DTN não há a presença de ACKs com relação ao envio de mensagens.

Figura 2.9: *Watchdog* gerando falso positivo.



Fonte: adaptado de (Marti et al., 2000).

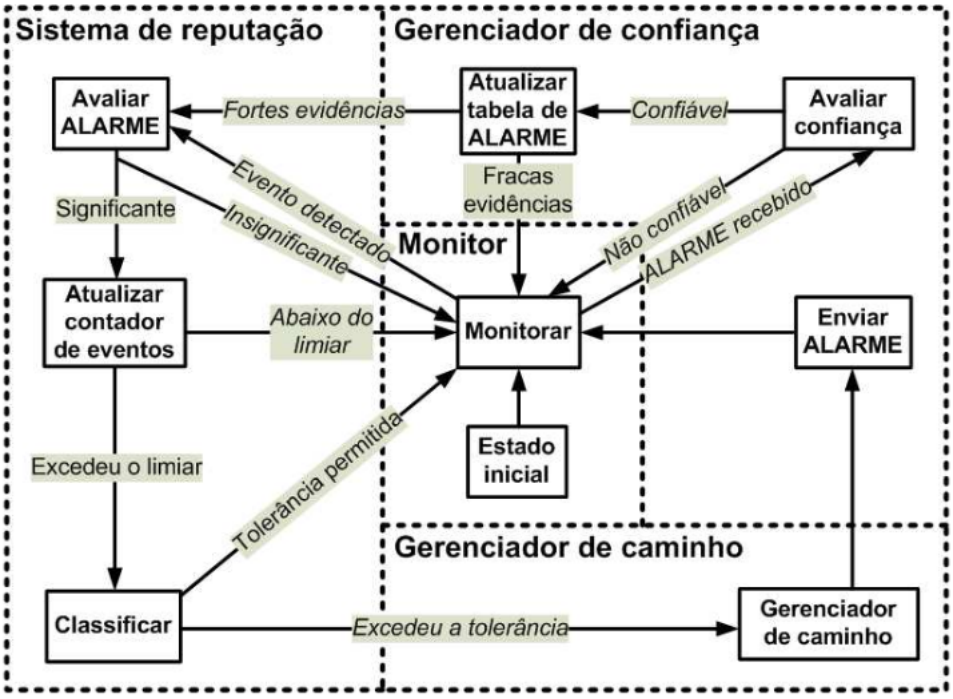
Em (Buchegger & Le Boudec, 2002) é apresentado o protocolo CONFIDANT. Esse protocolo utiliza a teoria de Dawkins (Dawkins, 1976), sobre altruísmo em ambientes ecológicos para criar um modelo que torna o comportamento egoísta pouco atrativo a rede. o protocolo CONFIDANT consiste de quatro componentes principais: o monitor, modelo de reputação, gerência de rota e o gerência de confiança que são representadas em detalhes na Figura 2.10. O primeiro componente consistem em analisar o canal dos nós da vizinhança (primeiro salto) e verificar um determinado comportamento como não repassar a mensagem. Quando um evento suspeito é detectado, a informação é passada para o modelo de reputação que reavalia a reputação do vizinho. Estas informações são utilizadas posteriormente pelos gerentes para montagem de um sistema inteligente de roteamento e gerência de confiança. Aqui Buchegger *et al.* constroem a reputação através de um modelo bayesiano conhecido como distribuição beta[5]. Cada resposta no monitoramento é do tipo contato egoísta e contato não egoísta, que é utilizado para alimentar a distribuição beta. Contudo, alguns estudos apontam que o uso de distribuições baseadas em proporções podem gerar vícios de aprendizagem (Cochran, 1954). Um exemplo que podemos escrever é o caso em que um nó que contribui muito com a rede, que, após certo tempo, é difícil e mais demorado detectar ele como egoísta quando seu comportamento muda.

Em DTN, um problema primordial é a conexão intermitente. O protocolo CONFIDANT, contudo, assume que é possível ouvir o canal de comunicação com a vizinhança até o momento que possa haver um repasse (segundo salto) na mensagem. Esse fator torna o sistema de monitoramento de vizinhança pouco preciso em relação a detecções dado a propriedade de conexão oportunista em DTN. Isto leva a perda de precisão dos componentes acima.

Michiarva e Molva (Michiardi & Molva, 2002) criaram um modelo de reputação colaborativo chamado CORE, um método baseado em *watchdogs*, contudo é comple-

[5]https://doi.org/10.1002/0471722227.ch16

Figura 2.10: Componentes do protocolo CONFIDANT.



Fonte: (Braga, 2008).

mentado por um modelo de reputação que realiza:

- **Reputação subjetiva**: reputação baseada nas próprias observações do sujeito. É calculado pela média ponderada de observações, atribuindo maior relevância a dados passado, com o intuito de minimizar falsos positivos esporádicos.
- **Reputação indireta**: avaliação feita pelo sujeito da comunicação através de sua vizinhança, que o comunica sobre a reputação subjetiva destes.
- **Reputação funcional**: Avaliação feita sob as reputações indiretas. Esse tipo de reputação serve para calcular um valor de reputação objetiva global considerando os valores observados.

Estes três itens são combinados em uma forma única para cálculo da reputação do conjunto vizinhança de um nó. CORE permite apenas informação positiva de segunda mão, isto é, por reputação indireta, que faz com que afirmações espúrias e nós mal intencionados incrementando reputações de outros prejudiquem o desempenho final deste método. Além disso, a falta de avaliação analítica ou simulada dificulta a compreensão deste método em ambientes DTN reais, embora a caracterízação do fluxo descrito pelos autores sirvam de referências até nos trabalhos atuais.

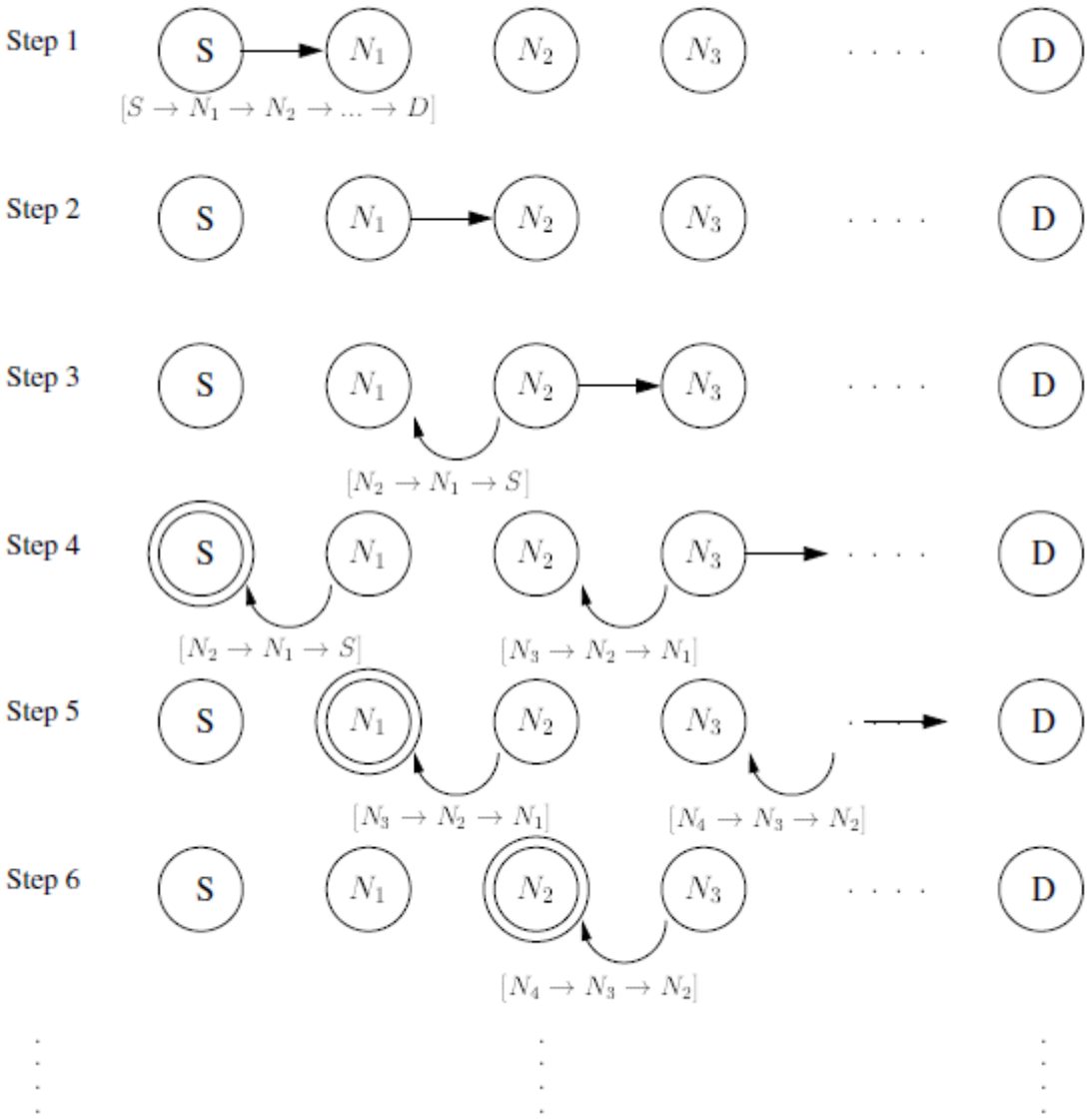
He *et al.* (He et al., 2004) propuseram o mecanismo SORI, no qual a reputação de um determinado nó é calculada através da probabilidade com que ele encaminha as mensagens de seus vizinhos. Assim, quando um nó recebe uma mensagem de um vizinho, ele a encaminha com a mesma probabilidade que o vizinho encaminha suas mensagens. Algumas premissas importantes de redes oportunistas como grupos sociais não são levadas em consideração, no entanto, assim nós menos populares ou nós com menos contatos terão dificuldades de terem suas mensagens encaminhadas em relação aos nós com mais popularidade ou contatos.

Em (Balakrishnan & Varshney, 2005; Kejun Liu et al., 2007) é proposto um esquema de verificação na camada de rede chamado TWOACK. Quando um nó repassa uma mensagem aguarda seu ACK e aguarda o ACK do próximo salto a partir do nó intermediário, assim garantido que o nó intermediário no primeiro salto repassou a mensagem adiante para no mínimo dois saltos a frente do nó de origem do dado. O funcionamento deste esquema é ilustrado na Figura 2.11 para um repasse de dados pela rota $[S \rightarrow N_1 \rightarrow N_2 \rightarrow \ldots \rightarrow D]$. Apesar de relevante, esse método enfrenta problema de congestão de ACKs na rede além de com a grande quantidade de ACKs gerar falso positivos devido a exclusão de IDs que não receberam TWOACK devido a *buffer overflow*. Outra dificuldade encontrada neste método que devemos ressaltar é a interferência dos padrões sociais de mobilidade que podem claramente afetar o funcionamento deste método caso este não possua um mecanismo de controle de congestionamento específico que leve em consideração padrões sociais.

Alguns mecanismos analíticos de deteção são mencionados em (Hernández-Orallo et al., 2013; Hernández-Orallo et al., 2015; Soares et al., 2015). Nestes trabalhos, a detecção é feita utilizando um *watchdog* analítico, isto é, cada nó possui um *watchdog* que funciona do seguinte modo: o *watchdog* de cada nó funciona com duas probabilidades, a primeira é a probabilidade de detecção P_d, que indica a probabilidade durante o contato de um nó identificar alguma informação se o nó pode ou não ser egoísta. Além disso há a probabilidade P_e que indica a eficiência do *watchdog* em detectar corretamente durante o contato. O exemplo na Figura 2.12 mostra um fluxo de identificação para o *watchdog* verificar quando o nó vizinho no contato é egoísta ou não. Primeiramente é verificada a probabilidade de detecção P_d. Caso a probabilidade seja menor que um limiar então o *watchdog* não tem parâmetros suficientes para realizar uma detecção, caso contrário é acionada a probabilidade P_e e, caso ela seja satisfeita ele identifica corretamente o nó como não egoísta, caso contrário identifica incorretamente como não egoísta. De modo similar a Figura 2.13 demonstra o fluxo para identificação de nós não egoístas.

Em (Li & Das, 2010), a estratégia proposta por Li e Das utiliza uma mensagem

Figura 2.11: Esquema de detecção de egoísmo TWOACK.

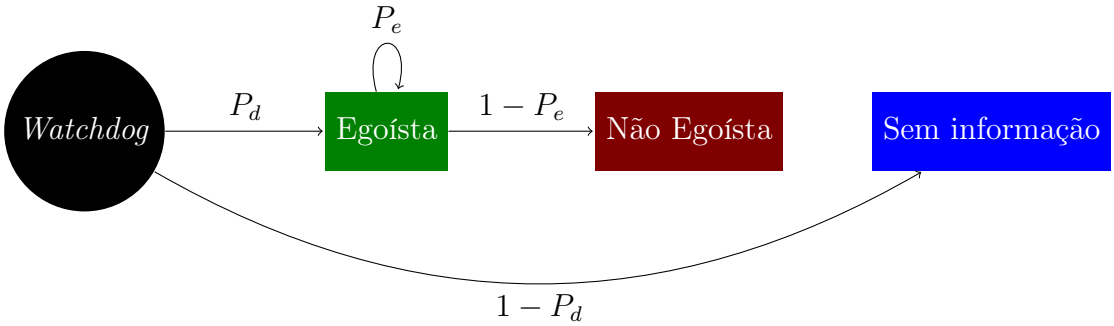


Fonte: (Balakrishnan & Varshney, 2005).

especial, chamada *Positive Feedback Message* (PFM) para ajudar a monitorar o comportamento colaborativo de repasse ou o descarte intencional de dados de um nó. Esse mecanismo é semelhante ao mecanismo de sinalização TWOACK, contudo o *feedback* é enviado em modo epidêmico e não apenas para o nó que repassou a mensagem. Com relação ao módulo de reputação, este estima o nível de confiança da comunicação entre cada nó i e j utilizando a distribuição de probabilidade $Beta(\alpha, \beta)$, onde α e β representam o número de *feedbacks* positivos e negativos, respectivamente, durante o período de tempo Δt.

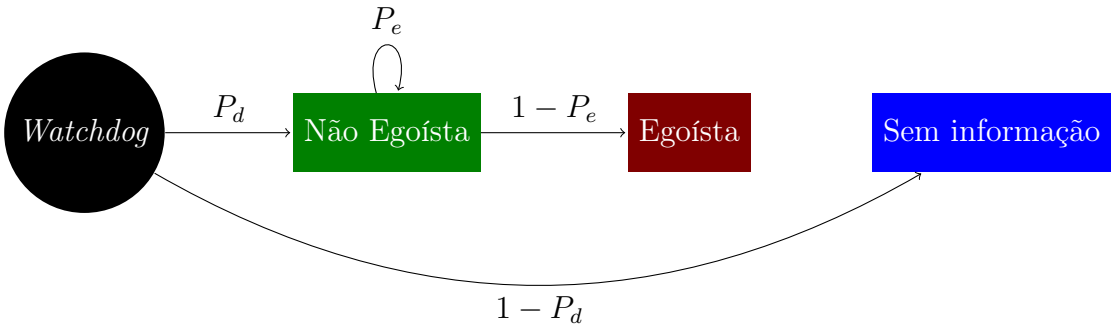
Esta inferência estatística, no entanto, não reflete com precisão o real desejo de

Figura 2.12: Fluxo de trabalho do *watchdog* para identificar um nó egoísta.



Fonte: o autor.

Figura 2.13: Fluxo de trabalho do *watchdog* para identificar um nó **não** egoísta.



Fonte: o autor.

repasse em redes de contatos oportunistas. Uma das principais contribuições deste trabalho é modelar este problema de acordo com a teoria de Dempster-Shafer[6] para quantificar a incerteza de variáveis aleatórias. Esta teoria permite combinar evidências (neste caso as observações) de diferentes fontes de dados ou de diferentes espaços temporais até alcançar um grau de crença que toma por considerações todas as evidências disponíveis.

Li et al. em (Li et al., 2010) proporam o método SSAR (*Social Selfishness Aware Routing*) para detecção de relação social aliado ao desejo de repasse e um novo método de repasse integrado. Esse protocolo usou priorização de mensagens como medida de desejo de repasse entre dois nós durante a comunicação. Assim temos que o *buffer* é organizado pelas prioridades das mensagens que é recalculada a cada salto da forma $p_i = p_{i-1} \times \omega$, onde ω é o desejo de um nó B repassar mensagens de uma comunicação com outro nó A.

Como percebe-se há uma linha tênue de métodos reputação aliada a interação

[6]http://aitopics.org/sites/default/files/classic/Buchanan/Buchanan15.pdf

social, padrão bastante estudado em seres humanos em diversos estudos. Com esta abordagem em questão Bigwood et al. em (Bigwood & Henderson, 2011) propuseram um mecanismo que perguntava a um nó sobre a qualidade de serviço de outros nós. Esse mecanismo verificava as comunicações do nó vizinho B e se esse fosse vizinho de outro nó ao qual o nó A repassou um determinado dado perguntava-lhe se a mensagem em questão havia sido repassada. Assim a cada valor negativo ou positivo era extraído um valor x do *rating* do nó de interesse.

À medida que muitos trabalhos focaram seus esforços em métodos de detecção de egoísmo para criar modelos de reputação, alguns trabalhos tem avaliado esse mecanismo sem a utilização de métodos de detecção (Dini & Lo Duca, 2010; Lu et al., 2010). Em (Dini & Lo Duca, 2010) a reputação dos nós é decrementada periódicamente e apenas é incrementada a reputação dos nós intermediários, isto é, os nós que fizeram a mensagem chegar até aquele ponto. Por exemplo suponho que o nó A tenha recebido uma mensagem M_k tal que o campo de rota seja $R = [S \leftarrow N_1 \leftarrow N_2 \leftarrow \ldots \leftarrow N_j]$, onde N_j representa o salto anterior antes de a mensagem chegar em A. Então todos os nós N_j tem sua reputação incrementada. Isso ocorre pois eles contribuiram na comunicação, isto é, no repasse ou, caso contrário, A não teria recebido a mensagem.

2.3 Técnicas de Agrupamento

Esta seção visa apresentar um conceito que será usado adiante no desenvolvimento deste trabalho. Como a classificação é feita baseada em agrupamento para escolha de nós egoístas e não egoístas baseado em reputação, foram pesquisadas algumas técnicas de agrupamento e ferramentas computacionais que poderiam separar as reputações em grupos de nós mediante valores de reputação. Neste trabalho destacam-se o estudo e utilização de algoritmos baseados em agrupamento, como o algoritmo *k-means* e o agrupamento hierárquico.

A extração de conhecimento a partir de informações é um estudo que pode ser aplicado para dividir um conjunto de informações em grupos (*clusters*) que tenham algum significado. Esse tipo de extração é conhecido como técnicas de agrupamento. As técnicas de agrupamento são aplicadas quando não existem classes previamente definidas, isto é, não existe uma base com histórico de informações que pode ser utilizada no decorrer de uma coleta de informações. Geralmente o uso dessa técnica é feito quando as amostras devem ser divididas em grupos com bases em suas características.

Técnicas de agrupamento têm sido amplamente utilizadas em muitos campos científicos (Gasch & Eisen, 2002; Holliday et al., 2004; Jain, 2010). Nas áreas de enge-

nharia, essas técnicas foram utilizadas para reconhecimento biométrico, fala, análise de sinais, remoção de ruídos em comunicações de rádio, entre outros campos de estudo. Também têm sido aplicadas em áreas da biologia como identificação de funções de genes, na química para caracterizar estruturas de componentes químicos e até mesmo na psicologia, para análise de perfis criminais.

Chamamos de grupo, todo conjunto de amostras que possui alguma semelhança de informações entre si. Logo, temos que semelhança é definido como o espaço métrico utilizado para definir o quão próximas estão as amostras ou o quão dissimilares são as amostras.

A similaridade é calculada através de funções geralmente no intervalo $[0, 1]$ ou $[0, \infty)$, onde 0 representa que objetos são extremamente semelhantes, enquanto 1 (ou ∞) representa a máxima diferença entre estes. As funções de similaridade mais utilizadas na literatura são apresentadas na Tabela 2.3

Tabela 2.3: Medidas de similaridade mais utilizadas em técnicas de agrupamento.

Distância	Função	Detalhes
Euclidiana	$D_{ij} = \sqrt{\sum_{l=1}^{d}(x_{il} - x_{jl})^2}$	Medida de semelhança mais utilizada
Manhattan	$D_{ij} = \sum_{l=1}^{d} \lvert x_{il} - x_{jl} \rvert$	Distância baseada em blocos quadriculares no plano $\mathbb{R}^2$
Cossenos	$S_{ij} = \cos\alpha = \frac{x_i^T x_j}{\lvert\lvert x_i \rvert\rvert \ \lvert\lvert x_j \rvert\rvert}$	Baseado na comparação de objetos com atributos contínuos. Amplamente usado em agrupamento de documentos

Desse modo, é dito que agrupamento representam técnicas de aprendizagem de máquina diferentes das clássicas classificações, pois são métodos de aprendizagem que buscam realizar associações ao invés de atribuir alguém a uma classe previamente conhecida. A esta metodologia chamamos de aprendizagem de máquina não supervisionada.

Assim, a aprendizagem de máquina pode ser dividida em dois modos: supervisionada e não supervisionada. Na aprendizagem supervisionada há a figura de professor, pelo qual apresenta conhecimento do ambiente através de um conjunto de dados de entrada, enquanto que, na aprendizagem não supervisionada, o objetivo é a identificação de padrões ou tendências que auxiliem no entendimento dos dados sem conhecimento prévio de informações que possam auxiliar (de Souto et al., 2003).

Entre as técnicas de agrupamento mais difundidas na literatura estão: os agrupamentos hierárquicos e *k-means* (Berkhin, 2006). Dado que o agrupamento realizado

neste trabalho é feito sobre um conjunto de dados 1-D (quando há apenas um atributo), também ponderamos sobre as técnicas: estimadores de densidade *kernel* (*kernel density estimator*) e otimização de jenks.

2.3.1 Agrupamento Hierárquico

Os agrupamentos hierárquicos relacionam suas amostras através de uma representação de árvore, também conhecida como dendograma, cujos comprimentos dos ramos refletem os grau de similaridade entre as amostras.

Esses relacionamentos são úteis pois eles podem representar graus variados de similaridade, além de requererem poucas suposições sobre a natureza dos dados (Jain & Dubes, 1988).

A construção do agrupamento hierárquico é feito da seguinte forma: inicialmente cada objeto pertence a um grupo único formado por ele mesmo; a cada interação, um grupo funde ao grupo mais próximo através das medidas de similaridade. Esse processo ocorre até que todos os pontos formem apenas um grupo (a hierarquia mais alta).

A vantagem no uso de agrupamento hierárquico se dá no fato de compreender a formação de grupos com pouca ou maior especialidade. Além disso, serve para evitar múltiplas execuções de agrupamentos, assim o primeiro agrupamento serve de base para decisões futuras. No entanto, isso torna o modelo sensível a ruídos e comportamentos dinâmicos. Assim, implicando que agrupamentos errados feitos inicialmente podem não ser corrigidos ao longo do tempo de observação dos objetos.

2.3.2 *K-means*

Diferentemente do agrupamento hierárquico, no qual o agrupamento funde ou particiona os grupos a cada iteração, no *K-means* o agrupamento é particionado em exatos K grupos, não havendo quaisquer hierarquia entre eles. Dado que o problema abordado neste trabalho não visa quaisquer hierarquia entre os nós baseado no valor de reputação, essa é uma estratégia que pode ser considerada mais atrativa.

Assim, no *k-means*, o objetivo traçado é encontrar um agrupamento tal que os grupos sejam tão homogêneos quanto possível e os objetos em cada grupo estejam bem distintos de objetos em outros grupos (Jain & Dubes, 1988).

Para isto, o *k-means* posiciona de forma aleatória K pontos centrais, chamados centroides, dentro espaço $\mathbb{R}^c$, onde c representa a dimensão utilizada e definida pelo número de atributos c utilizados nas amostras. Assim, cada amostra a é associada ao grupo mais próximo através da distância entre a e o centroide $C_i \forall i \in K$. Desse modo,

cada centroide é recalculado, a cada nova amostra que entra no grupo. A iteração do *k-means* termina quando o processo de atualização dos centroides para, significando assim que nenhuma nova associação pode ser feita, e os K grupos formados representam as similaridades mais próximas quanto possíveis.

Uma importante questão relacionada ao agrupamento realizado pelo *k-means* é a escolha do número de grupos K. Desde que K pode ser uma representação subjetiva, isto é, baseado na descrição do problema, a divisão em K grupos é naturalmente conhecida. Um exemplo pode ser dado utilizando este trabalho, no qual o problema remete à detecção de nós egoístas. Assim, é natural que $K = 2$ seja a opção mais atrativa, representando um grupo dos nós não egoístas e o grupo dos nós egoístas após a classificação.

Quanto a robustez, vale ressaltar que o *k-means* é sensível a ruídos e a presença de *outliers*. Como o cálculo dos centróides incluem todos os pontos, logo há a inclusão dos chamados *outliers*, pontos que são erroneamente agrupados. Contudo, algumas estratégias na literatura visaram o reconhecimento de *outliers* e a diminuição do impacto causados por este. O algoritmo PAM (Kaufman & Rousseeuw, 2008) utiliza o conceito de medóide, que é um ponto real do conjunto de dados que tenha a menor distância média para todos os outros pontos grupo. Assim, eliminando possíveis *outliers* do agrupamento ou não utilizando estes para alterar os valores do centróides.

Desde que em redes oportunistas, técnicas de detecção lidam com possíveis detecções erradas (*outliers*) em alguns instantes, então um mecanismo de agrupamento que possa leva-los em consideração é de suma importância.

2.3.3 Estimadores de Densidade *Kernel*

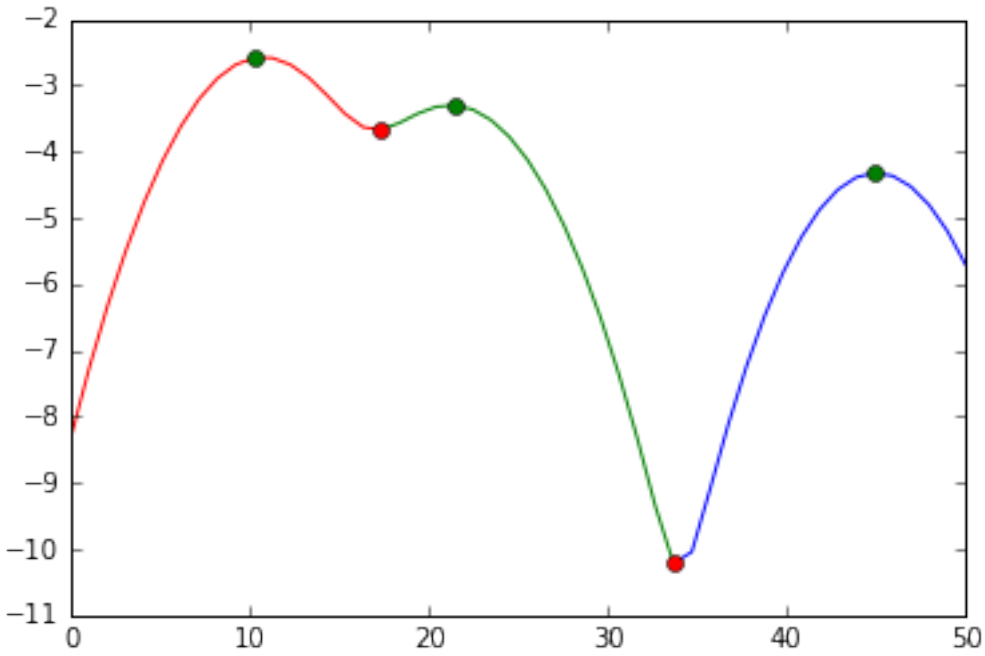
Os estimadores de densidade *kernel* são técnicas não paramétricas de estimar a função de probabilidade de uma variável aleatória (Silverman, 2018). Assim, esse método é considerado fundamental onde deseja-se encontrar inferências sobre uma determinada população.

De modo formal, suponha um conjunto $(x_1, x_2, \ldots, x_n)$ seja uma amostra de densidade f conhecida. A intenção é estimar o comportamento da função f. O estimador de densidade *kernel* é dado por:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K_{bw}(x - x_i) = \frac{1}{n.bw}\sum_{i=1}^{n} K(\frac{x - x_i}{bw}) \tag{2.1}$$

onde K é o *kernel* e bw é a largura de banda. O K_{bw} é o *kernel* atribuído por $K_{bw}(x) = \frac{1}{bw} * K(\frac{x}{bw})$.

Figura 2.14: Exemplo de encontro de mínimos locais para $k = 3$ grupos.



Fonte: imagem retirada de um exemplo disponível em https://bit.ly/2wRlU27.

Embora não seja uma técnica de agrupamento propriamente dita, uma forma de utiliza-lo como técnica de agrupamento é através da realização da criação do estimador *kernel* e realizar o cálculo para encontro dos k (número de grupos) maiores pontos locais separados por $k-1$ pontos de menor densidade, também chamados de mínimos locais. Um exemplo deste formalismo é descrito pela Figura 2.14.

Assim, o fluxo para o cálculo de agrupamento usando o estimador ocorre através dos seguintes passos: normalização e ordenação dos dados, computação de densidades, encontrar os k máximos locais e encontrar os $k-1$ mínimos locais e separação dos agrupamentos.

Mesmo apresentando um promissor método para clusterização, a estimação da função de densidade *kernel* depende de parametrizações como a largura de banda bw, que podem modificar o comportamento da função $f(x)$. Assim, a aplicação dessa técnica pode apresentar dificuldades de estimação em um ambiente de tempo real devido a escolha de bandas por cenário aplicado

2.3.4 Otimização de Jenks

O método de otimização de Jenks é um método de agrupamento construído para criar o arranjo de informações em k grupos com o objetivo de minimizar o desvio padrão

médio de cada classe enquanto maximiza o desvio padrão de cada grupo em relação aos outros grupos (Jenks, 1967). Logo, esse método busca diminuir a variância dentro dos grupos e maximizar entre grupos.

O processo de classificação é iniciado pela ordenação das amostras. Em seguida é construído o histograma de frequências, que auxilia na construção de possíveis agrupamentos. O cálculo dos limites das classes se baseia na estimativa do índice denominado melhor ajuste de variância (GVF – *Godness of Fitness*). Esse índice é utilizado para quantificar a qualidade da distribuição dos elementos nas classes de acordo com a similaridade entre as observações (Slocum & Egbert, 1993).

A principal vantagem desse método é a simplicidade para informações de uma dimensão. Esse método se assemelha à técnica *k-means* no modo como as amostras são atribuídas nos k grupos. No entanto, Jenks é apropriados para amostras unidimensionais ordenada, contudo seu processo pode ser desacelerado devido à otimização na busca pela variância mínima dentro dos grupos, enquanto que o *k-means* pode ser considerado um caso geral. A otimização de Jenks também é referenciada como *k-means* aplicados a dados univariados. Embora este trabalho tenha usado uma base unidimensional, optamos por utilizar o *k-means*, pois torna nosso modelo proposto maleável a inserção de novas informações nas amostras.

Capítulo 3

Arquitetura do Sistema

Neste capítulo, nós descrevemos a arquitetura de rede utilizada neste trabalho, o modelo desenvolvido para a detecção e o modelo de reputação dos nós na rede oportunista utilizada para avaliação de desempenho. Identificamos os principais desafios relacionados a este trabalho e damos uma visão geral do modelo de rede utilizado, seguido por uma descrição detalhada do modelo de reputação utilizado para classificar nós egoístas em redes oportunistas. Além disso, descrevemos em detalhes o modelo utilizado para realizar a classificação dos nós baseado no valor da reputação através das técnicas de agrupamento.

3.1 Visão Geral e Desafios

Modelos de mobilidade em redes compostas por humanos, mudanças de topologia, desconexões e dificuldades na detecção de nós egoístas representam muitos desafios para a concepção e implementação de métodos para classificação de nós egoístas em redes oportunistas. A estratégia proposta neste trabalho é um modelo de reputação para redes oportunistas. Seu funcionamento básico está ligado à troca de contatos para elaboração de um ranqueamento à respeito do comportamento dos nós da rede. Nosso mecanismo contabiliza resultados de detecção e utiliza informações locais e oriundas de outros nós para elaboração do ranqueamento. Além disso, cada nó é responsável pelo gerenciamento de modelo de reputação, visto que nestes tipos de rede não há autoridade central.

Esta seção descreve detalhadamente uma visão geral dos desafios que motivaram as premissas utilizadas na modelagem e implementação do modelo proposto neste trabalho.

3.1.1 Modelo de Decisão Distribuída

A conectividade limitada e a natureza fragmentada das redes oportunistas significam que não é possível obter nem manter um nível global de conhecimento da rede ou utilizar técnicas baseadas em reconhecimento para decisões globais na rede. Em vez do modelo tradicional de redes com infraestrutura fixa, cada dispositivo da rede oportunista deve agir de forma independente e baseado apenas no conhecimento limitado obtido a partir das observações locais. Assim a distribuição de tomada de decisões neste trabalho podem ser divididas em duas partes:

- **Escopo de informações**: o modelo utiliza informações de primeira mão (experiência local) e informações compartilhadas por outros nós da rede.
- **Gerenciamento de reputação descentralizado**: o modelo de reputação de um nó é gerenciado localmente e independe de outros nós da rede. Dessa forma, cada nó não requisita informações dos outros nós sobre sua tomada de decisão.

Adicionalmente, quando os nós da rede compartilham informações na rede a taxa de sucesso no compartilhamento da informação é mais alto, principalmente no âmbito de detecção de egoísmo em redes DTN (Hernández-Orallo et al., 2015; Silva et al., 2016). O uso de informações indiretas, isto é, oriundas de outros nós, trás alguns questionamentos sobre a segurança de informações compartilhadas. Contudo, isto não é profundamente estudado neste trabalho e consideramos que o compartilhamento de informações não sofre de alterações, injeção de informações falsas na rede nem falhas de segurança severas.

3.1.2 Grupos Sociais em Redes Oportunistas

A conectividade esporádica da rede formam ilhas de comunicações, nos quais os nós se comunicam em grandes grupos baseados em suas relações sociais. Desse modo, os nós oportunistas podem estar presentes em diversas comunidades ou em poucas comunidades, tornando o conhecimento global da rede com relação a um objeto de interesse uma tarefa mais complexa.

Certo trabalhos (Marti et al., 2000; Michiardi & Molva, 2002; Buchegger & Le Boudec, 2002) estudaram técnicas de detecção de egoísmo que utilizavam determinados limiares para identificar nós egoístas baseados na reputação destes nós. Contudo, muitos destes trabalhos foram desenvolvidos em redes cuja a mobilidade era totalmente aleatória, no entanto, devido aos padrões de mobilidade e grupos sociais em redes oportunistas e DTN a concentração de informação pode ser menor em determinados

pontos da rede. Além disso, conhecer o comportamento de nós cuja popularidade ou centralidade seja menor na rede pode requerer mais tempo e esforço.

Assim, é crucial a identificação de comportamento mesmo quando haja pouca informação. Neste trabalho, ilustramos alternativas possíveis para uma detecção mais eficiente mesmo para nós cuja reputação é pouco conhecida. Para isto, a utilização de limiares verificada em trabalhos anteriores foi removida e utilizamos técnicas de aprendizagem de máquina com o intuito de detectar mais rapidamente o comportamento dos nós. Um estudo mais amplo das vantagens e desvantagens de tal abordagem são discutidos na seção de resultados.

3.1.3 Comportamento Dinâmico

Chen e Chen afirmam sobre o comportamento de um nó com relação à utilização de recursos (Chen & Chen, 2007):

- O comportamento de um nó pode mudar ao longo do tempo, enquanto a qualidade do recurso não.
- O critério de avaliação dos nós apresenta diferentes aspectos em relação à capacidade desse nós.

Podemos perceber que o comportamento de um nó pode mudar. Assim como em outros modelos de reputação existentes em outros campos de estudo, o comportamento está intrinsecamente ligado ao desejo de participação do agente no sistema. Mais especificamente em redes DTN e redes oportunistas, o desejo de participação está ligado ao uso de recursos e desejo de repasse de suas mensagens na rede. Quando um recurso vital de um nó está escasso ou o nó deseja assumir o risco de não ter suas mensagens repassadas na rede, o comportamento egoísta se torna uma saída plausível. Quando a quantidade de recursos são alteradas ou o desejo que suas mensagens sejam repassadas, o comportamento pode ser modificado afim de atingir os objetivos individuais de cada nó.

Desse modo, as seguintes características dinâmicas foram consideradas na construção do nosso modelo:

- **Precisão**: o valor de reputação deve representar o comportamento com o maior nível de confiança possível.
- **Rápida convergência**: o valor de reputação calculado deve se adaptar rapidamente a mudança de comportamento do nó. Um nó não pode se beneficiar da

boa reputação e, em seguida, tentar se beneficiar dessa boa reputação quando agir de modo egoísta.

- **Adaptação dinâmica aos participantes**: nós podem entrar e sair no sistema. O modelo deve lidar com essa dinâmica, em vez de situações pré-determinadas.
- **Escalabilidade**: o modelo deve lidar com sistemas largos sem perda de desempenho o quanto possível.

3.2 Modelo de Rede

Nesta seção, focamos no processo de modelagem da rede e como ela foi utilizada no âmbito deste trabalho.

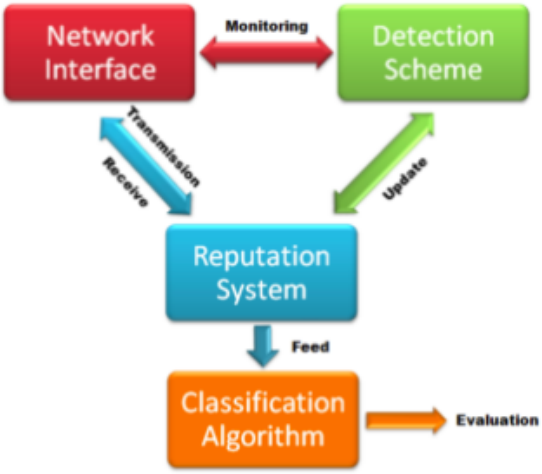
Para realização da arquitetura da solução proposta baseada em reputação, o modelo de rede utilizado é constituído a partir de uma simulação de rede na qual o conjunto de conexões é modelado como um grafo $G(V, E)$, tal que V é o conjunto de vértices ou nós da rede, E o conjunto de contatos oportunistas e o contato é definido pelo par ordenado (u, v), $u, v \in V$, no qual $d_{(u,v)}$ é a relação que define que o nó u tem uma demanda de envio de mensagens para o nó v, também chamada de oportunidade de contato.

Cada nó $u \in V$ é um dispositivo munido de *buffer*, e uma arquitetura responsável por gerenciar repasses e controle de *buffer*. Também consideramos que cada dispositivo é igual ao outro, assim todos os nós da rede são idênticos estruturalmente entre si no início da simulação.

Assim, o processo de comunicação na rede em cada nó é definido conforme a Figura 3.1. Cada nó possui uma interface de rede que realiza o monitoramento dos repasses na rede, através de um método de detecção, comumente chamado de *watchdogs*, um modelo de reputação que é atualizado conforme as detecções ocorrem durante os contatos oportunistas.

Adicionalmente, este trabalho propõe a inserção de uma camada de classificação. Essa camada é responsável por dizer o comportamento de um nó dada sua reputação. Diferentemente de outros trabalhos na literatura (Marti et al., 2000; Buchegger & Le Boudec, 2002), que utilizaram limiares para classificação, este trabalho apresenta esse processo de classificação via agrupamento integrado ao modelo de reputação pela primeira vez, pelo nosso conhecimento.

Figura 3.1: Arquitetura de um nó da rede utilizando mecanismo de detecção e modelo de reputação.



Fonte: o autor em (Soares et al., 2015).

3.3 Mecanismo de Detecção de Nós Egoístas

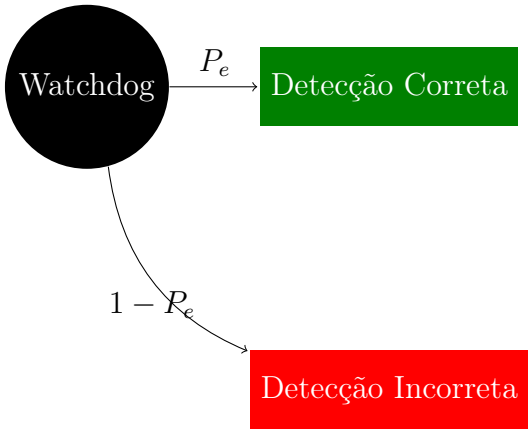
O foco deste trabalho é o modelo de reputação que será explicado mais adiante. No entanto, para isto é preciso um mecanismo que detecte comportamento egoísta em redes DTN.

Neste trabalho a detecção de nós egoístas foi feita de modo analítico, conforme já utilizado em trabalhos anteriores (Hernández-Orallo et al., 2013; Soares et al., 2015). Cada nó possui um *watchdog* que funciona com uma probabilidade de eficiência na detecção, P_e, cujo funcionamento é descrito pela definição abaixo:

Definição 3.3.1. P_e é a probabilidade de eficiência do *watchdog* durante o monitoramento. Dado que *watchdogs* estão sujeitos a falhas por motivos de intermitência no canal de comunicação, controle de congestionamento (quando o número de pacotes de controle do *watchdog* cresce rapidamente) do *watchdog*, gerenciamento de *buffer* (método que realiza a gerência de mensagens armazenadas de modo persistente no nó intermediário), entre outros. O valor de P_e é definido no início da simulação

Conforme definimos tal probabilidade, o fluxo de como a detecção funciona é demonstrado na Figura 3.2, no qual em verde temos o estado em que a informação detectada pelo *watchdog* é correta, enquanto em vermelho a detecção incorreta que ocorre com probabilidade $1 - P_e$ em relação ao número de detecções.

Figura 3.2: Fluxo de trabalho do *watchdog* para identificar um nó egoísta.



Fonte: o autor.

Ressaltamos que a escolha de não implementarmos um *watchdog* foi devido à carência de estudos que pesquisaram profundamente o funcionamento do *watchdog* em redes oportunistas com modelos de mobilidade baseados ou extraídos de ambiente reais bem como a eficiência destas estruturas em ambientes simulados. Desse modo, para podermos ter uma maior flexibilidade e controle do nível de eficiência do *watchdog*, optamos por este modelo analítico de projetar um *watchdog*.

3.4 Modelo de Reputação dos Nós

Neste trabalho propomos um mecanismo que pontua o grau de colaboração dos nós da rede baseado em histórico de detecção. Os métodos de detecção como *watchdogs* estão sujeitos à erros na fase de detecção devido à erros no canal de comunicação ou informações erradas propagadas por nós vizinhos. Por isso, um método baseado em pontuação foi escolhido. Outro motivo é que em alguns ambientes podemos encontrar egoísmo em nível social e não individual, isto é, com graus de egoísmo que variam conforme as relações sociais entre os membros da rede. Portanto, procuramos desenvolver um método que pudesse suprir esses dois casos e utilizar o fator de colaboração para calcular mais rápido o índice de detecção associado com a precisão de detecção.

Nesta seção apresentamos o modelo utilizado neste trabalho para realizar o ranqueamento. Primeiramente representamos o problema de cooperação como uma disputa durante o contato, isto é, cada contato resulta em uma competição com saída de positivo ou negativo para cooperação do segundo contato. Depois demonstramos uma forma de realizar essa competição de maneira distribuída.

É fato que em alguns tipos de jogos como pôquer ou xadrez, a habilidade do jogador é mensurável (Charness et al., 2005). Podemos imaginar um jogo de acerto do valor de um dado. Nesse jogo você ganha se acertar o número obtido após o dado ser jogado. Depois de algumas tentativas, haverá acertos e principalmente erros. Um conhecido axioma de probabilidade nos dirá que nesse jogo, quando jogarmos um número quase infinito de vezes teremos:

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \tag{3.1}$$

No qual $P(E)$ é a probabilidade de ocorrência de um evento E quando este evento ocorre infinitas vezes. Como possuímos um dado com probabilidade 1/6, então haverá a probabilidade de ganharmos uma a cada seis jogos quando a ocorrência desse evento é levada ao infinito em número de ocorrências. Contudo, imagine um jogo em que a medida de probabilidade de um evento não é uniformemente distribuída. Uma maneira de visualizarmos seria imaginar um jogador comum contra um grande campeão de xadrez ou damas. É trivial analisar que intuitivamente a probabilidade de um vencer o outro não é 1/2. A esse tipo de comportamento, chamamos de índice de reputação.

Visualize um jogo onde há dois participantes e esses participantes gostariam de saber a possibilidade de um terceiro jogador ganhar de ambos em partidas apenas entre dois participantes. Se ao menos um deles conhecer o histórico do terceiro jogador, será possível que eles façam estimativas sobre a probabilidade de vitória desse terceiro jogador. A esse passo, no qual usuários compartilham informações passadas sobre um evento ou agente chamamos de passo colaborativo.

Atribuindo o cenário como sendo uma rede oportunista formada por nós móveis, reduzimos este problema a um problema de teoria dos jogos, a cada contato $C(i, j)$ em que entre dois nós i e j, a reputação entre ambos é calculada. Nesse caso, estamos interessados em dois cenários de contato: contato entre um nó cooperativo e outro egoísta, e o contato entre dois nós cooperativos, este último sendo tratado adiante.

No primeiro cenário, $C(i, j)$ é um contato entre um nó cooperativo i e um vizinho egoísta j. Cada um dos nós tem uma probabilidade P_e de detectar corretamente o comportamento do outro nó, como realizado em (Hernández-Orallo et al., 2013) para detectar nós egoístas.

Redes oportunistas possuem a particularidade de conexão intermitente e altas taxas de erros, assim a fase de detecção do nó egoísta pode gerar falsos positivos, quando um nó é detectado como egoísta e não é egoísta, por exemplo. Com o intuito de minimizar falsos positivos e assumindo que com tempo podemos conhecer um nó egoísta através de um número de contatos que segue $\lim_{n \to \infty} N(C(i, j))$, tal que $N(k)$

representa o número de contatos $C(i, j)$ entre dois nós durante o tempo de vida de ambos, elaboramos um modelo de atualização de reputação a partir da interação entre pares de contatos.

Assumindo que no instante anterior à comunicação do nó cooperativo i que $V(i) \cap V(j) = \varnothing$, tal que $V(i)$ representa o conjunto vizinhança conhecido do nó i, então atribuímos uma reputação inicial neutra R para o nó j. Logo, o nó i tem interesses de repasse da sua lista de mensagens e precisa saber se o nó j é egoísta. Dado que o nó i tem uma probabilidade de detecção e essa probabilidade é positiva para possível nó egoísta naquele instante, então atualizamos a reputação do vizinho da seguinte forma (um detalhamento maior é dado na sequência do texto):

1. Nó verifica a reputação conhecida de todos os seus vizinhos.

2. Nó calcula a probabilidade de nó j ser mais cooperativo que os vizinhos já conhecidos de i como se fosse um jogo $P_{cf}(k, j) \mid k \in (V(i) - \{j\})$, no qual $J(k, j)$ representa um confrontamento entre j e os outros k vizinhos conhecidos do nó i. O cálculo $P_{cf}(k, j)$ é feito com a reputação do nó j e a reputação média dos vizinhos definida pela equação 3.3.

3. O nó atualiza a reputação do outro nó através da equação 3.4, no qual K é um fator de peso para atualização e $D(i)$ é o resultado da detecção de comportamento sobre outro nó.

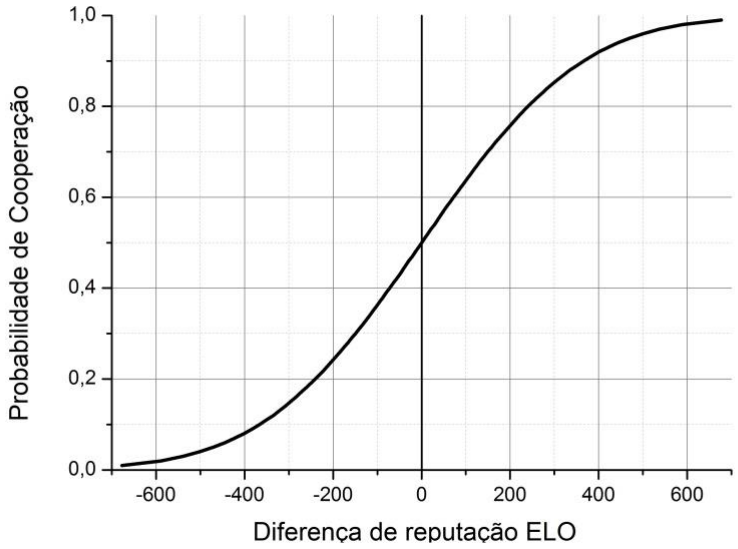
Nesse trabalho modelamos a probabilidade de um nó ser mais cooperativo que outro dado a reputação atual de ambos através da função de força relativa abaixo:

$$P_{cf}(k, j) = \frac{1}{1 + 10^{\frac{(R_j - R_k)}{400}}} \tag{3.2}$$

A função acima é também conhecida matematicamente como ELO *rating*[1] $P_{cf}(k, j) : (\mathbb{N}, \mathbb{N}) \rightarrow \mathbb{R}, P_{cf}(i, j) \in (0, 1)$, uma função logística de inferência estatística criada para modelar o grau de habilidade entre jogadores de xadrez e comumente usada para predizer resultados de partidas de futebol, desempenho de usuários em competições de informática e sistemas de reputação em redes sociais. Essa função possui uma importante propriedade de aprendizagem durante o tempo de observação do sistema, no qual o estado inicial de crescimento é aproximadamente exponencial até que durante um período de reconhecimento são feitas análises no sistema e essa função, então, cresce de forma mais lenta até alcançar um ponto de maturidade. A variação de

[1] https://www.gautamnarula.com/rating/

Figura 3.3: Diferença de valores de reputação e probabilidade de cooperação futura para o nó k.



Fonte: o autor.

probabilidade de cooperação dado a diferença de reputação (*rating*) entre dois usuários utilizando essa função é mostrada na Figura 3.3.

Para todos os vizinhos de i menos j temos que j foi detectado não cooperativo naquele instante de tempo então podemos aumentar a reputação dos outros vizinhos já conhecidos a partir de suas reputações já conhecidas. Para realizar essa atualização de modo descentralizado primeiro definimos as equações a seguir:

$$MR(i) = \frac{\sum_{\forall k \in V(i)-\{j\}} R_k}{|V(i)| - 1} \tag{3.3}$$

$$R_j^{'} = R_j + K(D(j) - P_{cf}(j,k)) \tag{3.4}$$

$$D(j) = \begin{cases} 1 & se \text{ i} = \text{ colaborativo} \\ 0 & \text{caso oposto} \end{cases} \tag{3.5}$$

A equação 3.3 representa a reputação média dos vizinhos conhecido pelo nó i no instante em que um outro j nó é detectado na comunicação.

A equação 3.4 representa a atualização da reputação para o nó comunicante. Para realizar isto, supomos que um contato representa uma comunicação entre nó

colaborativo e nó egoísta, o primeiro cenário. Se um nó é detectado egoísta então ele precisa ter sua reputação, no ponto de vista do nó i, diminuída. Então para o nó j temos que $D(j) = 0$. Posteriormente, o processo de atualização se divide em duas etapas: a atualização da reputação dos nós conhecidos de i e a atualização da reputação do nó j, o nó em contato atual com i.

A primeira etapa consiste em avaliar os nós vizinhos já conhecidos, tal que haja um ganho de reputação quando um nó j é detectado egoísta através da equação 3.4. A segunda etapa da atualização consiste em reavaliar o nó detectado egoísta j. Como avaliar o nó j individualmente com todos os vizinhos de i implica em uma grande variabilidade na reavaliação de j então optamos por utilizar uma medida de média aritmética para todas as reputações conhecidas entre os vizinhos de i, representado pela equação 3.3. Logo, temos que $D(j) = 0$ e, assim, $P_{cf}(j, k)$ é calculado com referência à média de reputação na equação 3.3, e então atualiza sua reputação conforme 3.4. O procedimento atual para essa comunicação é apresentado no algoritmo 1.

Algoritmo 1: Algoritmo para atualizar reputação do nó j

Data: $i, < j, R_j >$
Result: `// reputação atualizada`
$K = Fator_{inicial}$ **if** $j \notin V(i)$ **then**
    $R_j = Reputacao_{inicial}$
    $V(i).add(< j, R_j >)$
else
    `// reputação atual de` j `conhecida por` i
    $R_j = V(i)[j]$

`// Nó` j `foi detectado como egoísta por` i `com probabilidade` P_e `de acertar`
if *isSelfishAction(j)* **then**
    foreach $\forall k \in \{V(i) - \{j\}\}$ **do**
        $D(k) = 1$
        $R_k = R_k + K(D(k) - P_{cf}(k, j))$ `//` P_{cf} `conforme eq. 3.2`

    $< m, R_{medio} > = < m, MR(i) >$ `// Conforme eq. 3.3`
    $D(j) = 0$
    $R_j = R_j + K(D(j) - P_{cf}(j, m))$

Dado que aqui, a probabilidade de um nó ser egoísta é imutável ou sofre pouca variação (se i é frequentemente detectado como egoísta, então ele não mudará radicalmente seu comportamento para altruísta durante os contatos com um mesmo nó).

Portanto, os nós podem aprender o grau de colaboração dos seus vizinhos quando um ponto de maturidade é alcançado.

Portanto, temos a situação que i detectou j como colaborativo e este por sua vez requisita informações sobre os vizinhos de j, afim de montar uma rede colaborativa de conhecimento. Nesta etapa é realizado o passo colaborativo, no qual os nós compartilham as informações sobre seu conjunto vizinhança agregando informações. Alguns trabalhos propõem nesta etapa utilizar um fator constante w para atualizar o modelo de reputação quando a informação é passada de segunda mão, isto é, atribuindo mais peso a detecções próprias que a detecções de outros nós da rede (Berger, 1985). No entanto, devido a premissa de que um nó não é malicioso ao mesmo tempo que é colaborativo então utilizamos a média aritmética para equilibrar medidas que ambos tenham sobre um mesmo vizinho. Assim definimos formalmente o cálculo como mostrado em 3.6, no qual $\alpha = \beta = 0.5$, representando a média aritmética simples.

$$\begin{gathered} R_{(k,i)} \in \{R_i \mid i \in V(i)\} \\ R_{(k,j)} \in \{R_j \mid j \in V(j)\} \end{gathered} \tag{3.6}$$

$$\forall k \in (V(i) \cup V(j)) \rightarrow R_k = \alpha R_{(k,i)} + \beta R_{(k,j)}$$

Durante um contato entre os nós A e B se o nó B possuir informações sobre um nó que A não tem informações prévias, então este adquire a reputação que B conhece sobre este nó. Ressaltamos aqui que a premissa é que nenhum nó cria ou compartilha informações falsas na rede.

Assim, cada nó i tem uma tabela de reputações $R_k \forall k \in V(i)$, na qual cada entrada R_k corresponde a um *rating* sobre um nó vizinho k. Aplicado os *ratings* aos nós vizinhos no tempo t, então o nó i pode utilizar essa tabela para alimentar seu sistema classificador, descrito detalhadamente na seção 3.5.

3.5 Classificação do Comportamento

Trabalhos prévios sobre modelagem por reputação utilizam normalmente limiares para classificação dos nós em egoístas ou não egoístas. Quando um valor de reputação ultrapassa um limiar, este passa a sinalizá-lo como egoísta. Devido aos padrões de mobilidade e padrões de interação social, a definição de limiares pode ser uma tarefa complexa dado que em alguns casos a informação é mais concentrada em partes da rede e a definição de limiares menos brandos pode dificultar na detecção de nós egoístas, por exemplo, em nós que têm um círculo social mais fraco.

Este trabalho inova ao utilizar métodos classificadores no âmbito de classificação de egoísmo baseado nos valores de reputação. A premissa que partimos aqui é de que dado um conjunto de valores de reputações R_i, $\forall i \in V$ podemos separá-los em k classes tais que as separações dos valores R_i em cada classe é suficiente para gerar uma classificação eficaz e possivelmente livre de falsas classificações. Para isto utilizamos uma abordagem bem conhecida na área de aprendizagem de máquina e mineração de dados, a técnica de agrupamento.

O propósito do agrupamento é a definição de grupos intrínsecos em um conjunto de dados que não possuem rótulo, de modo que cada objeto agrupado em uma classe possuam semelhanças perante algum critério prévio. Uma das grandes vantagens dessa técnica é também a classificação em grupos sem o conhecimento prévio de uma base de conhecimento, estratégia conhecida na área de aprendizagem de máquina como aprendizagem de máquina não-supervisionada (Jain & Dubes, 1988). As aplicações de agrupamento também se estendem a diversos campos de estudos como bioinformática (Golub et al., 1999), análise de dados web (Vakali & Pallis, 2007) e reconhecimento de escrita manual (Kato & Nemoto, 1996).

Para sumarizar, o agrupamento é um modo de aglomerar dados usando alguma medida de similaridade específica, tentando rotulá-los afim de que objetos do mesmo grupo tenham características mais semelhantes do que objetos pertencentes a outros grupos.

Em vista dessa definição, algumas considerações importantes são levantadas no âmbito da realização de agrupamentos, tais como:

- Como medir a similaridade entre os objetos?
- Como formar os agrupamentos?
- Quantos grupos formar?
- Como validar os agrupamentos formados?

A medição de similaridade entre objetos está relacionada aos dados disponíveis para realizar a similaridade como dados numéricos ou categóricos. Para isso existem três formas de medição da proximidade: medidas correlacionais, medidas baseadas em distância e medidas de associação (Hair et al., 2005).

Quanto a forma de realização de agrupamento há as formas hierárquicas e não hierárquicas, no qual as técnicas hierárquicas montam uma estrutura no qual cada grupo é ligado a outro numa forma de árvore enquanto nas formas não hierárquicas

os grupos são divididos sem que haja qualquer ligação hierárquica entre elas (Jain & Dubes, 1988; Hair et al., 2005).

Sendo assim, a definição formal de agrupamentos é definida da seguinte maneira: dado um conjunto de entrada $X = \{x_1, \ldots, x_N\}$, em que $x_i = (x_{i1}, x_{i2}, \ldots, x_{id}) \in \mathbb{R}^d$ onde cada medida x_{ij} é chamada de característica ou atributo (Jain & Dubes, 1988):

1. Um agrupamento não hierárquico tenta buscar uma partição de X, $C = \{C_1, \cdots, C_k\}$ com $k \leq N$, tal que,
   - $C_i \neq \varnothing, i = 1, \ldots, k$
   - $\bigcup_{i=1}^{k} C_i = X$
2. Um agrupamento hierárquico tenta construir uma partição de X, $C = \{C_1, \cdots, C_k\}$ a partir de uma estrutura aninhada, $H = \{H_1, \ldots, H_Q\}$ com $Q \leq N$, tal que,
   - $C_i \in H_m, C_j \in H_l$
   - $C_i \subset C_j$ ou $C_i \cap C_j = \varnothing, \forall i, j$ com $i \neq j$

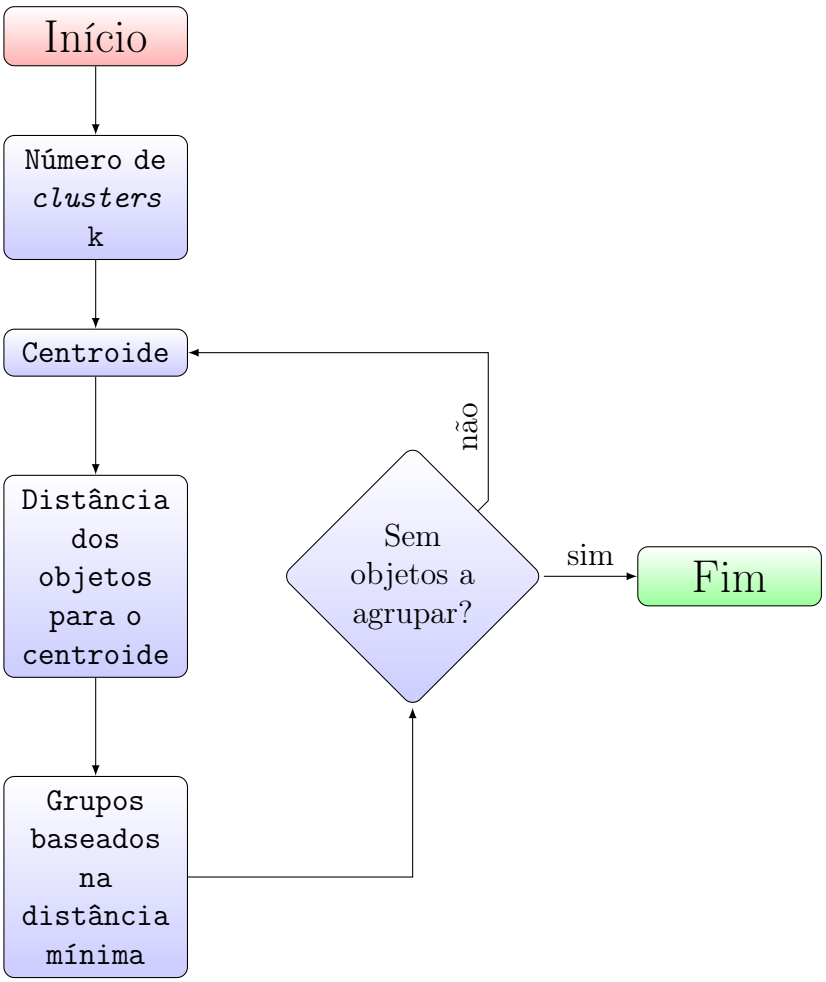
Neste trabalho, devido à caracterização individual do egoísmo, isto é, nenhum nó é egoísta porque outro nó é egoísta por exemplo, então escolhemos utilizar o agrupamento não hierárquico para realização do agrupamento.

Um clássico algoritmo não hierárquico de agrupamento presente na literatura é o algoritmo *k-means*. A ideia base do *k-means* é formar grupos a partir de um ponto central do grupo chamado de centroide. Assim a ideia base do *k-means*, conforme definido em (Hartigan & Wong, 1979) é ilustrado pelo fluxograma da Figura 3.4.

Para realizar a tarefa de agrupamento, o *k-means* busca minimizar o erro quadrático dos pontos espaciais em relação aos centros de seus grupos. Quando os atributos dos objetos estão bem separados, o erro quadrático se torna mínimo. Assim quando um conjunto de atributos não possua boa separação para definição de classes, o erro quadrático se torna maximizado.

Quanto ao número de agrupamentos, embora haja técnicas e heurísticas que indicam o valor k ideal de grupos dados os atributos para todos os objetos presentes, escolhemos utilizar $k = 2$ grupos, para representar o grupo dos nós egoístas e o grupo dos nós não egoístas.

Figura 3.4: Fluxograma do algoritmo *k-means*.



Fonte: o autor, baseado na descrição dos autores em (Hartigan & Wong, 1979).

Capítulo 4

Implementação e Avaliação

Neste capítulo apresentamos a avaliação de desempenho do modelo proposto de modo a identificar nós egoístas em redes DTN. A técnica escolhida para avaliação foi simulação, devido a grade complexidade de realização de experimentos em ambientes reais. Contudo, a mobilidade utilizada foi extraída de ambientes reais de modo a aumentar a similaridade com ambientes DTN reais.

4.1 Implementação

Na avaliação deste trabalho foi utilizado o simulador *open source* The ONE (*Opportunistic Network Environment*) (Keränen et al., 2009), desenvolvido em Java, ele simula os principais aspectos de redes DTN como transferência em custódia, roteamento e algoritmos de gerenciamento de *buffer*. Também permite a modificar o código fonte e parametrização de configuração dos nós da rede bem como a mobilidade dos nós. Suas principais características são:

- The ONE oferece suporte a modelos de mobilidade extraídos de ambientes reais (chamados de *trace*), troca de mensagens, roteamento DTN, consumo de energia e protocolos de aplicação.
- The ONE é projetado de modo modular, permitindo extensão de todas as funções usando interfaces bem definidas.

Para implementação do método de detecção de nós egoístas criamos uma classe *watchdog* que tem relação com a classe `DTNHost.java`, já implementada no simulador. O funcionamento dessa classe é explicado com maiores detalhes na seção 3.3 deste trabalho.

Além disso, foi implementada a tabela de reputação e os métodos de atualização e cálculos de reputação utilizando a classe `ActiveRouter.java` do simulador The ONE. O funcionamento detalhado desse mecanismo é descrito com detalhes na seção 3.4 deste trabalho.

Para agrupamento das amostras e posterior classificação delas foi utilizada a biblioteca `SimpleKMeans`, presente no software Weka (Hall et al., 2009), que foi integrado ao simulador The ONE.

4.2 Caracterização do Cenário

Simulações de ambientes moldados para redes oportunistas pressupõem a utilização de mobilidade de dispositivos carregados por usuários em ambientes reais. Assim, a utilização de mobilidade baseadas em *trace* de contatos é de suma importância. Neste trabalho, escolhemos a utilização do *trace* Haggle-Infocom05 (Chaintreau et al., 2007). Esee *trace* foi colhido durante uma conferência científica, na qual alguns participantes, munidos de sensores tiveram seus contatos e mobilidade extraídos durante o período da conferência. A principal razão pela escolha desse *trace* é a densidade de contatos entre os usuários deste modelo, tornando mais estável a interação social entre eles. Assim, garantimos que a arquitetura proposta não seja influenciada pela escolha de nós extremamente populares como egoístas em uma simulação, enquanto em outra simulação podemos escolher nós com muitos poucos contatos como egoístas, alterando significantemente a estabilidade da análise de desempenho. Detalhes extras sobre esse modelo são mostrados na Tabela 4.1.

Tabela 4.1: Características do *trace* Infocom5.

Característica	Infocom5
Número de nós	41
Duração (dias)	3
Tipo de dispositivo	iMote
Número de contatos	22459
Contatos por minuto	4,902
Tempo médio de contato (s)	231,755
Média da maior componente detectada	5,051
Grau médio dos nós	0,477

Utilizamos três períodos de tempo para realização de análises: 24, 48 e 72 horas de observação. Assim, nosso modelo foi avaliado nessas três durações. A principal motivação para isto foi a observação desse *trace* ser baseado do modelo diário de um

ambiente real, no qual uma parte do dia ocorrem os contatos, enquanto que outra parte de um dia os nós não realizam contatos, simulando o ambiente em condições noturnas.

Os parâmetros de detecção foram ajustados entre [75%, 95%] para a probabilidade de eficiência na detecção, também chamada de P_e. A escolha dos nós egoístas foi feita de modo aleatório, entre três valores de porcentagem de nós da rede como egoístas, 10%, 25% e 50% dos nós da rede como egoístas. Dado que nossas simulações pretendiam compreender a reputação dos nós sob condições de contatos, exclusivamente, mensagens transmitidas durante o tempo de vida da rede, roteamento e tamanho de *buffer* de dados não são considerados relevantes para uma análise mais profunda de parametrização.

4.3 Métricas Avaliadas

Para avaliarmos nosso modelo de reputação, trabalhamos em duas linhas: a caracterização da reputação e a classificação dos nós em egoístas dado o valor de reputação.

Para verificar a caracterização da reputação utilizamos um modelo bem conhecido de caracterização de variáveis aleatórias, o estimador de densidade *kernel*. Basicamente, consiste na realização de uma inferência sobre as amostras através de um método não paramétrico. Dado que não conhecemos previamente o modo como a reputação será caracterizada no espaço amostra, os estimadores *kernel* (Scott, 2008) surgem como uma boa solução para compreender tal comportamento e estabelecer hipóteses. Ao adotar uma hipótese relativamente fraca de que a verdadeira densidade é suave, permitimos que os dados contem ao analisador qual é seu verdadeiro padrão de comportamento (Bessegato et al., 2006).

Para avaliar a classificação utilizamos: o nível de separação no agrupamento e a taxa de acertos.

Com relação à validação da separação da reputação em agrupamentos, a tarefa básica nesta etapa é validar a eficiência do agrupamento em relação ao conjunto de dados utilizados, neste caso, a reputação. Basicamente existem duas formas de validação de agrupamentos: baseados em conhecimento prévio e aqueles baseados apenas nas informações intrínsecas dos dados (Jain & Dubes, 1988). Devido ao fato que neste trabalho assumimos que não há conhecimento prévio dos dados, a classificação foi feita baseada apenas nas informações disponíveis, também conhecida como classificação não supervisionada, na área de aprendizagem de máquina.

A avaliação do desempenho da classificação por agrupamento foi feita através do índice *Silhouette* (Rousseeuw, 1987). O objetivo deste índice é analisar o quanto um objeto é semelhante a outro dentro do seu próprio *cluster*, assim avaliando o quão bem

separado os k grupos estão. O índice *Silhouette* varia entre $[-1, 1]$. Valores ótimos do índice são próximos de 1, e valores acima de $0,7$ são indicadores para um bom agrupamento (Rousseeuw, 1987). Enquanto valores próximos a -1 são indicadores de agrupamentos com verossimilhança inversa e indicadores próximos de 0 indicam que o agrupamento não encontrou um padrão bom para realização do agrupamento.

Formalmente o cálculo do índice de *Silhouette* é feito conforme a equação 4.1, onde $a(i)$ é a média da distância de i com relação aos dados de todos os outros objetos dentro do mesmo *cluster* e $b(i)$ é a média da distância de i em relação a todos os outros objetos dos outros *clusters* presentes.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{4.1}$$

Assim a média global do índice de *Silhouette* para medição da qualidade do agrupamento é medido pela equação 4.2, na qual k representa o número de agrupamentos realizados.

$$\overline{silhouette} = \frac{\sum_{i=1}^{k} s(i)}{k} \tag{4.2}$$

Além disso, avaliamos a acurácia do agrupamento. Para isto, analisamos como cada nó i foi classificado via agrupamento e qual o real comportamento deste nó i, isto é se ele era egoísta ou não durante a simulação. Assim analisamos a taxa de verdadeiro positivo (TPR – *True Positive Rate*), também conhecida como precisão[1], que corresponde a taxa dos nós classificados corretamente como nós egoístas e é formalmente definido na Equação 4.3. Neste caso, consideramos que *true positive* (TP) é a principal classificação no contexto de comportamento egoísta em redes oportunistas.

$$TPR = \frac{TP}{TP + FP} \tag{4.3}$$

A principal justificativa dessa avaliação é verificar como os *outliers* afetam na classificação. Para todos os efeitos, um *outlier* neste trabalho é considerado todo nó cujas informações não são suficientes para determinar uma classe correta, tal que este nó possui características para estar em ambas as classes. Em outras palavras, os *outliers* podem induzir a análise errônea no momento da sinalização do *cluster*. Os motivos que levam a isto são os padrões de mobilidade e a probabilidade de eficiência na detecção do *watchdog* P_e.

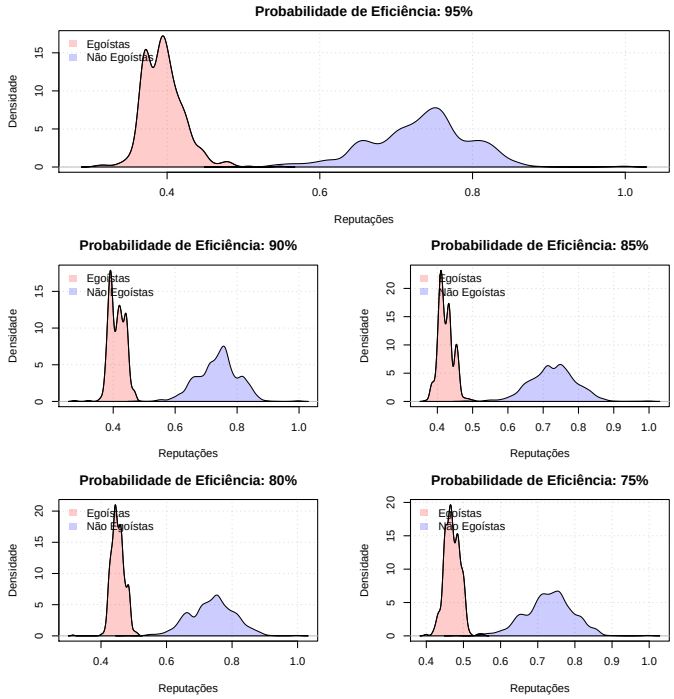
[1] https://pt.wikiversity.org/wiki/Observatório_de_dados/Precisão_e_revogação

4.4 Discussão dos Resultados

4.4.1 Análise do Modelo de Reputação

Para analisar nosso modelo de reputação, foram geradas as estimativas de densidade *kernel* das reputações dos nós egoístas e não egoístas em todos os nós. A razão disto é que ela é uma forma não paramétrica de estimar a função densidade de probabilidade de uma variável aleatória.

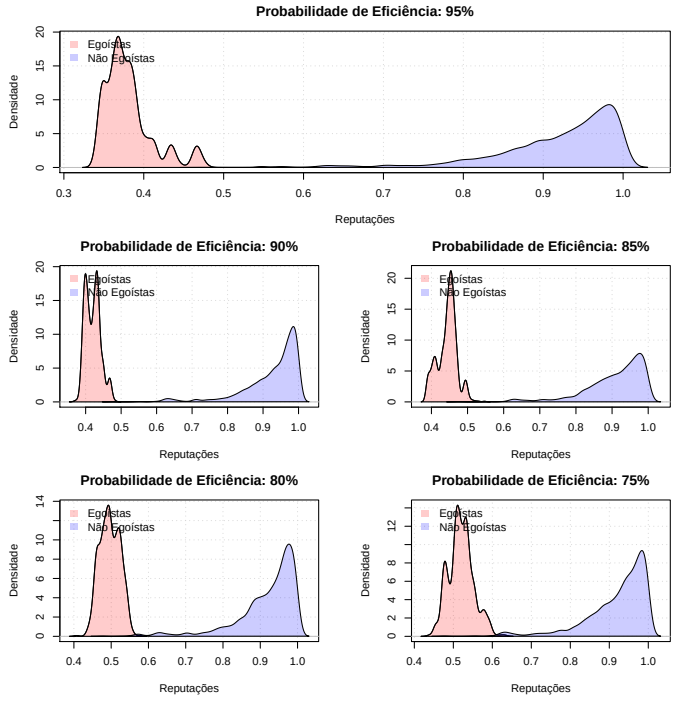
Figura 4.1: Estimativa de densidade *Kernel* das reputações dos nós com 24 horas e 10% dos nós egoístas.



Nas Figuras 4.1, 4.2 e 4.3, é mostrado a distribuição *kernel* das reputações para 24, 48 e 72 horas de simulação, respectivamente, e 10% dos nós como egoístas. Neste experimento avaliamos a probabilidade de eficiência na detecção do *watchdog* P_e com os valores $75\%, 80\%, 85\%, 90\%, 95\%$. O eixo x representa os valores de reputação dos nós, o eixo y representa a densidade das amostras coletadas naquela média de valores de reputação. Em rosa é dado os valores de reputação coletados de todos os nós egoístas (pela tabela de reputação dos nós que coletaram informações deles). Em azul é dado os valores de reputação conhecido pelos vizinhos para os nós não egoístas. Para todos os efeitos, o gráfico maior em cada figura apresenta a estimação *kernel* para $P_e = 95\%$.

Os valores de reputação se referem ao valor de reputação de um nó na tabela de reputação de todos os vizinhos que possuem alguma informação sobre o nó. As

Figura 4.2: Estimativa de densidade *Kernel* das reputações dos nós com 48 horas e 10% dos nós egoístas.

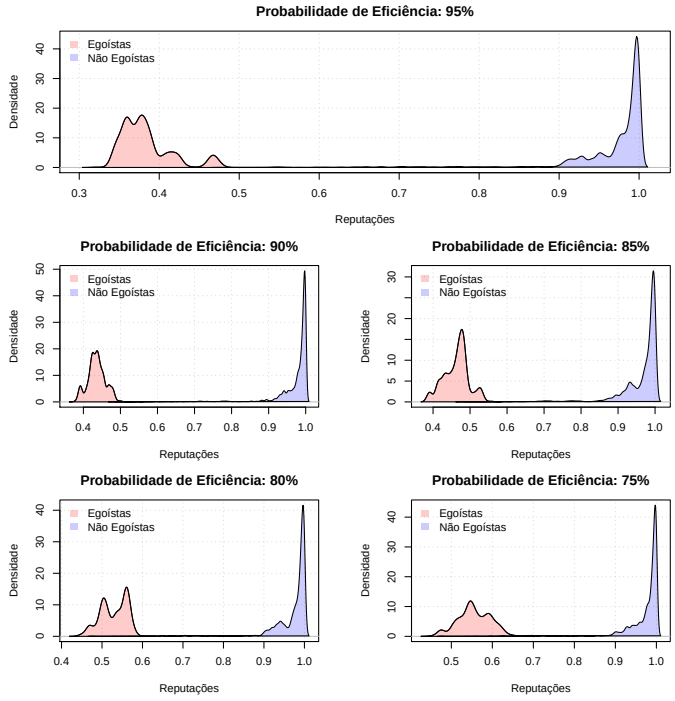


amostras são originadas a partir de um conjunto de simulações do mesmo cenário, no qual a variação é feita na escolha dos nós egoístas, que são escolhidos aleatoriamente. Para exemplificar o funcionamento dos gráficos, suponha que um nó u é conhecido pelos nós $\{v, w, j\}$ com reputação $R_u = \{0,80, 0,75, 0,90\}$ conhecida pelos nós $\{v, w, j\}$, respectivamente, então os três valores de reputação serão compilados no gráfico. Caso o nó u seja nó não egoísta, então os três valores serão computados na distribuição *kernel* em azul, caso contrário, na distribuição rosa.

Como pode ser verificado, há uma rápida convergência das reputações (para o valor 1 dos nós não egoístas e 0 para nós egoístas), principalmente para os nós não egoístas a cada passo de 24 horas. Com 24 horas de monitoramento, como demonstrado na Figura 4.1, a média de reputação dos nós egoístas variou de 0,47 para 75% de eficiência na detecção para 0,39 para 95% de eficiência na detecção como pode ser visto no gráfico. Enquanto isso, a média de reputação para os nós não egoístas quase não sofreu variação, ficando em torno do valor de ranqueamento $\simeq$0,73 para todos os valores P_e no método de detecção.

Um aspecto que vale destacar na Figura 4.1 é a distribuição esparsa dos valores de reputação dos nós não egoístas. Em modelos de reputação previamente descritos na literatura, o uso de limiares seria aplicado. Contudo, a porcentagem de nós que seriam classificados (com valores de reputação maiores que um limiar) seria baixa (abaixo dos

Figura 4.3: Estimativa de densidade *Kernel* das reputações dos nós com 72 horas e 10% dos nós egoístas.



50% no caso da Figura 4.1). Desse modo podemos concluir, que através dos estimadores *kernel*, além da boa separação entre valores de reputação, a fase de classificação, conforme será descrito adiante será feito em 100% da totalidade das amostras, demonstrando que nosso método pode classificar uma quantidade muito maior de amostras, quando comparado com trabalhos anteriores.

Com 48 horas de monitoramento, a média de reputação dos nós egoístas variou entre 0,5 até 0,37 para 75% de eficiência na detecção e 95% de eficiência na detecção, respectivamente. Para 72 horas de monitoramento, a variação foi de 0,54 até 0,36 para os valores de reputação dos nós egoístas. Quando analisamos a reputação dos nós não egoístas, no entanto, o resultado apresentado apresenta valores de reputação promissores, acima de 0,9 para todas as amostras avaliadas.

Neste caso percebemos que, quando o número de nós egoístas é baixo, o erro associado no mecanismo de detecção afeta muito na hora de gerar a reputação dos nós egoístas da rede. A hipótese que assumimos para tal resultado é o baixo número de observações na rede. Assim, quando uma detecção errada acontece, a reputação de um nó aumenta. Contudo, se houverem poucos contatos com este nó no futuro, não haverá chances de detecta-lo corretamente, diminuindo as chances de a reputação destes nós convergirem para um valor menor.

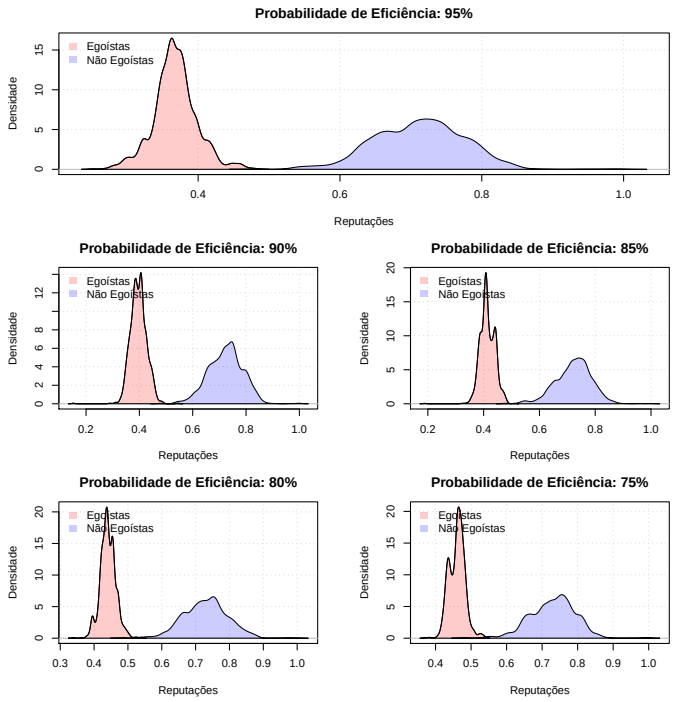
Além disso, analisamos o desvio padrão das amostras afim de verificar a dispersão

das reputações e, embora a convergência de reputação dos nós egoístas para 0,0 (menor valor possível de reputação e indicando o nível máximo de conhecimento sobre um nó egoísta) seja mais lenta do que nós não egoístas para 1,0 (maior valor possível de reputação e indicando o nível máximo de conhecimento sobre um nó não egoísta), o desvio padrão da reputação dos nós variou de 0,02 para 24 horas de observação até 0,03 para 72 horas de observação para os nós não egoístas, representando assim uma dispersão menor da reputação nos nós egoístas. Enquanto isso, para os nós egoístas o desvio padrão foi de ≃0,065 para 24 horas, ≃0,075 para 48 horas e ≃0,055 para 72 horas de observação, independente da probabilidade de eficiência na detecção.

Assim, podemos concluir que a convergência da reputação tende muito mais rápido para a quantidade máxima de nós na rede que possuem o mesmo comportamento. Como 90% dos nós da rede, não eram egoístas, a quantidade de observações trocadas era maior, tornando assim o conhecimento da rede sobre nós não egoístas maior.

Também analisamos a estimação da reputação quando 25% dos nós da rede eram egoístas. As Figuras 4.4, 4.5 e 4.6 apresentam os resultados para 25% dos nós da rede como nós egoístas após 24, 48 e 72 horas de observação.

Figura 4.4: Estimativa de densidade *Kernel* das reputações dos nós com 24 horas e 25% dos nós egoístas.



Nossa hipótese ao aumentar a quantidade de nós egoístas na rede era verificar se nosso modelo se manteria estável ao calcular as reputações. Com mais nós egoístas na rede, poderia haver o risco da reputação dos nós não egoístas não convergir tão

Figura 4.5: Estimativa de densidade *Kernel* das reputações dos nós com 48 horas e 25% dos nós egoístas.

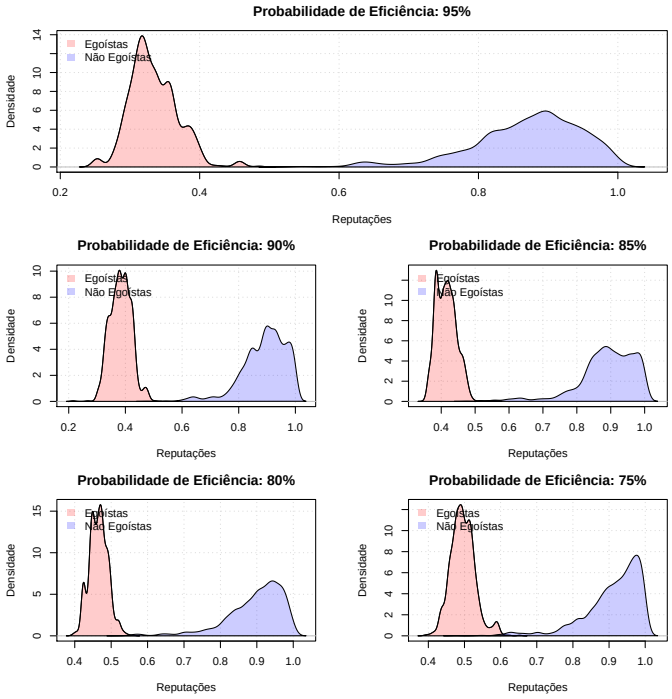
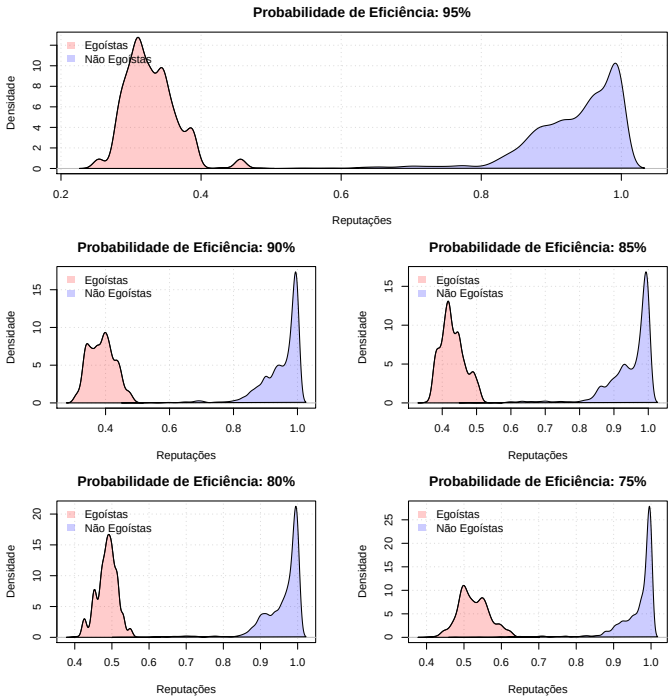


Figura 4.6: Estimativa de densidade *Kernel* das reputações dos nós com 72 horas e 25% dos nós egoístas.



rapidamente, tornando a separação de reputação dos nós egoístas e nós não egoístas uma tarefa mais árdua. Contudo, nossos resultados demonstram uma estabilidade ao apresentar resultados semelhantes aos encontrados para 10% dos nós egoístas. A diferença entre a média de reputação com relação do período de 24 horas e 48 horas, tanto para nós egoístas como para nós egoístas teve uma convergência mais lenta, próximo de 5% a 10%. Isso pode ser percebido ao compararmos visualmente as Figuras 4.4 e 4.5. Isto é a média de reputação converge para 0,0 (menor valor possível de reputação para nós egoístas) e 1,0 (maior valor possível de reputação para nós não egoístas) apenas entre 5% a 10% mais lento quando aumentamos de 10% para 25% de nós egoístas na rede.

Outra hipótese levantada foi o aumento da média dos valores de reputação dos nós egoístas quando a probabilidade de eficiência na detecção é menor, ou seja, uma diminuição na precisão da atribuição dos valores de reputação dos nós egoístas. Contudo, o aumento foi de apenas 5% em média. Isso significa que nosso modelo não reage bem na atribuição de valores de reputação de nós egoístas, quando a probabilidade de eficiência na detecção é mais baixa (75%). No entanto, devido a boa convergência dos valores de reputação dos nós não egoístas (próximos do valor máximo $1,0$), conseguimos verificar pelos gráficos a boa separação entre as reputações.

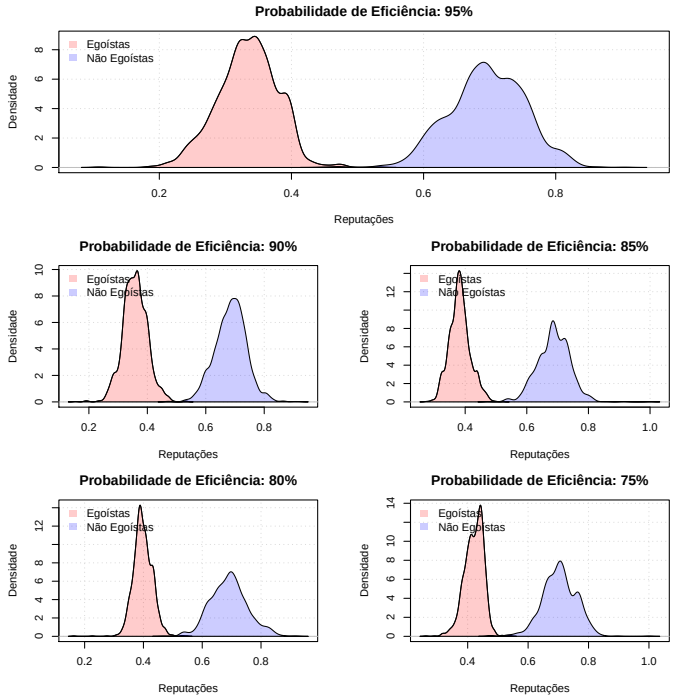
As Figuras 4.7, 4.8 e 4.9 apresentam os resultados da reputação média dos nós para 50% dos nós da rede egoístas.

Quando a quantidade de nós egoístas na rede incrementa muito (de 25% para 50%), percebemos uma significante mudança no modo com a reputação média dos nós se comporta.

Para os nós egoístas, observamos que a reputação média dos nós foi de 0,33, 0,26 e 0,23 para 24, 48 e 72 horas de observação com 95% de eficiência nas detecções. A média aumenta quando a taxa de acertos na detecção é de apenas 75%, com médias de 0,42, 0,40 e 0,40 para 24, 48 e 72 horas de observação. Além disso observamos um pequeno aumento no desvio padrão das amostras para nós egoístas, embora seja demonstrado pelos gráficos que a dispersão ainda é muito baixa mesmo com um tempo de observação de apenas 1 dia. Assim demonstramos que o modelo é estável e suficiente para identificar nós egoístas logo no primeiro dia de observação.

Para os nós não egoístas, observamos que a reputação média dos nós foi de 0,69, 0,82 e 0,86 para 24, 48 e 72 horas de observação com 95% de eficiência nas detecções. Diferentemente do caso em que os nós são egoístas, a eficiência na detecção menor causa menor impacto. Quando a probabilidade de eficiência é de 75% as médias foram de 0,70, 0,82 e 0,87 para 24, 48 e 72 horas de observação. A principal justificativa para isso é que quando dois nós não são egoístas se encontram e detectam o vizinho

Figura 4.7: Estimativa de densidade *Kernel* das reputações dos nós com 24 horas e 50% dos nós egoístas.



como não, a reputação deles é melhorada em cada nó, assim a informação dos nós não egoístas se propaga mais rapidamente que a dos nós egoístas. Embora isso pareça uma falha, a justificativa aqui é que nosso trabalho escolheu disseminar informações de nós não egoístas mais rápido que de nós egoístas. Dado que uma tomada de ação aqui poderia ser priorizar nós com maiores reputações para realizar repasses, por exemplo, então essa poderia ser uma boa abordagem a ser seguida.

4.4.2 Avaliação do Agrupamento

O agrupamento foi feita com o algoritmo k-means conforme descrito na Seção 3.5, cuja comentários sobre o método utilizado para avaliar o agrupamento são feitos também. Assim, avaliamos o coeficiente de silhueta de todas as clusterizações feitas. Os resultados são apresentados na Tabela 4.2.

Como podemos ver, o coeficiente de silhueta variou entre 0,60 e 0,62 para todos os resultados. Isso demonstra uma boa separação entre as reputações em cada *cluster*. Para fins de conhecimento, quanto mais perto de 1,0, o coeficiente de silhueta representa amostras bem separáveis. Um exemplo retirado do experimento com 10% de nós egoístas e $P_e = 75\%$ é apresentado na Figura 4.10, no qual o *cluster* 1 representa os nós egoístas e o *cluster* 2 representa os nós não egoístas. Como podemos observar,

Figura 4.8: Estimativa de densidade *Kernel* das reputações dos nós com 48 horas e 50% dos nós egoístas.

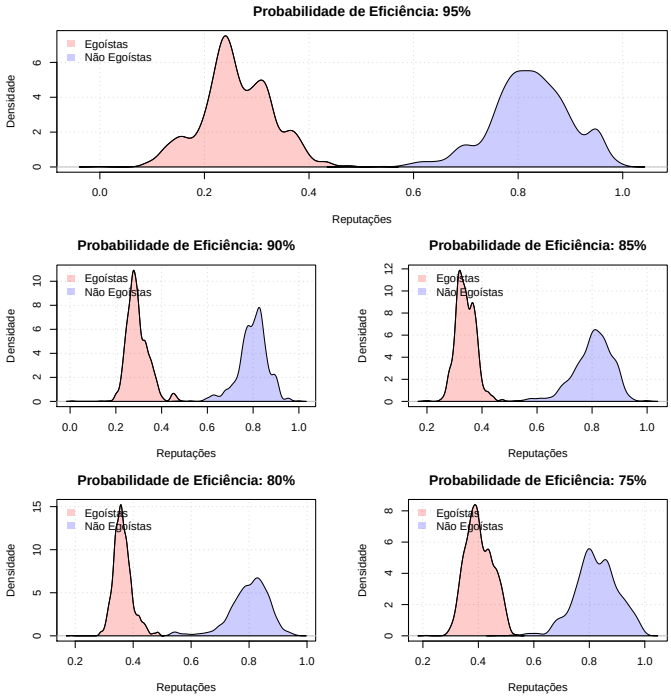


Figura 4.9: Estimativa de densidade *Kernel* das reputações dos nós com 72 horas e 50% dos nós egoístas.

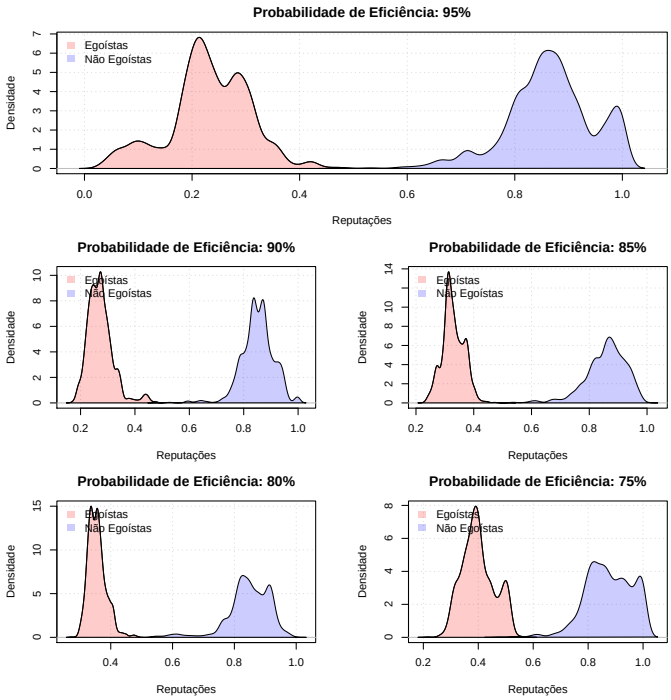


Tabela 4.2: Coeficientes silhuetas gerais em cada experimento.

		P_e				
% de nós egoístas	Tempo de simulação	95%	90%	85%	80%	75%
10% de nós egoístas	24 horas	0,6237	0,6226	0,6228	0,6230	0,6242
	48 horas	0,6215	0,6220	0,6221	0,6213	0,6223
	72 horas	0,6207	0,6203	0,6205	0,6201	0,6208
25% de nós egoístas	24 horas	0,6233	0,6232	0,6229	0,6219	0,6225
	48 horas	0,6212	0,6214	0,6221	0,6209	0,6209
	72 horas	0,6202	0,6198	0,6194	0,6200	0,6196
50% de nós egoístas	24 horas	0,6234	0,6248	0,6231	0,6240	0,6232
	48 horas	0,6212	0,6209	0,6206	0,6228	0,6212
	72 horas	0,6184	0,6203	0,6194	0,6205	0,6194

grande parte das amostras ficam bem separáveis em cada *cluster*, com coeficiente de silhueta acima da média. As amostras que ficam muito abaixo da média representam os resultados daqueles nós que pouco tem contato com os nós egoístas, assim, dificultando a atualização de suas respectivas reputações.

Tais resultados demonstram que o agrupamento mediante valores de reputação resulta em uma boa separação e se mantém estável mesmo quando a eficiência na detecção P_e é variada de 75% até 95%, o que nos faz crer que o modelo de agrupamento é estável em variados cenários, tornando independente do erro associado na detecção e da porcentagem de nós egoístas na rede.

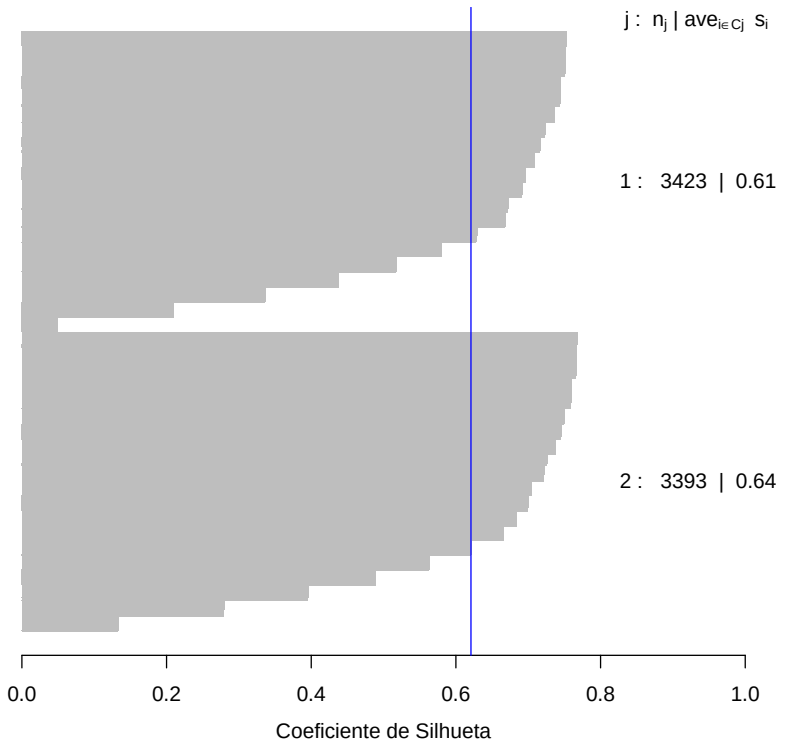
Em (Soares et al., 2015), a mesma técnica é aplicada em um outro cenário e apresenta coeficientes de silhueta em torno de 0,7. Esse trabalho corrobora que o agrupamento é uma metodologia que pode ser aplicada para o tratamento das reputações dos nós da rede.

4.4.3 Avaliação da Classificação

De modo a realizar um ambiente em que cada nó é classificado como egoísta baseado no valor da reputação, nós calculamos a TPR, taxa de verdadeiros positivos, conforme descrito na seção 4.3.

Primeiramente, avaliamos a TPR quando 10% dos nós da rede eram egoístas. A Figura 4.11 demonstra os resultados para $P_e \in 75\%, 80\%, 85\%, 90\%, 95\%$ e para tempos de observação de 24, 48 e 72 horas. Conforme podemos verificar no gráfico, o resultado da classificação quando o número de nós é baixo é mais inconsistente quando o tempo de monitoramento é mais baixo, contudo, conforme o tempo de monitoramento aumenta, a taxa de acerto chega a passar de 90% mesmo quando $P_e = 75\%$. Além

Figura 4.10: Coeficiente de silhueta das amostras no experimento com 10% de nós egoístas e $P_e = 75\%$.

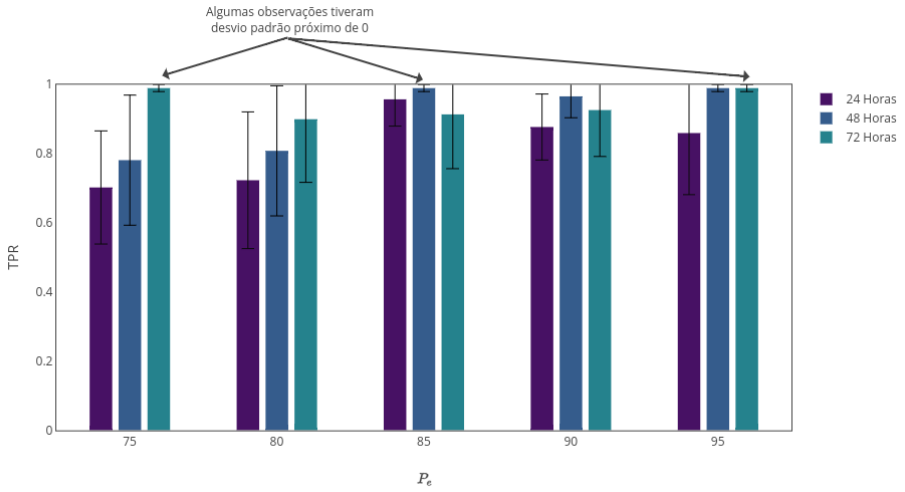


do mais detectamos que a variação é maior quando o número de nós egoístas é mais baixo. Dado que as informações dos nós egoístas se propagam mais lentamente quando o número de nós egoístas é baixo, a informação demora mais a ser conhecida na rede. Percebemos porém, que quando P_e é maior que 85% a taxa de acerto fica superior a 90% mesmo com o tempo de observação em torno de 24 horas.

Desse modo, percebemos que quando o número de nós egoístas é baixo, é coerente que a detecção seja produza resultados melhores, em torno de mais de 85% de detecções corretas, assim mantendo a TPR em um alto nível de aceitação e baixo nível de variação.

Conforme já havíamos percebido na seção 4.4.1, quando temos menos nós egoístas na rede, a convergência da reputação dos nós egoístas converge muito mais lentamente do que a reputação dos nós não egoístas. Quando analisamos o TPR para 25% e 50% dos nós egoístas, este comportamento ficou mais evidente, no qual TPR foi perto do valor perfeito de 100%. Assim, analisamos o TPR com tempos diferentes, analisando após 12, 18 e 24 horas de observação. A razão disto foi que em quase 100% das amostras, o TPR foi de 100% para 48 e 72 horas de observação.

Figura 4.11: Taxa de verdadeiros positivos (TPR) para 10% dos nós da rede como egoístas.



Na Figura 4.12 é apresentado o resultado para 25% dos nós sendo egoístas na rede. Percebemos que o TPR foi maior que 97% em todos casos e alcançando a 100% de acerto através do desvio padrão das médias. Quando o aumento do número de nós egoístas aumenta, além da variação diminuir na simulação, aumentamos significativamente a taxa de acertos dos nós que são egoístas.

Ressaltamos neste ponto que não necessariamente todos os nós da rede conhecem todos os nós egoístas da rede em 12, 18 e 24 horas de simulação, mas aqueles que possuem um algum valor atribuído de reputação para os nós egoístas, estes acertam em classificar os nós egoístas em mais de 97% dos casos.

Enquanto isso, a Figura 4.13 apresenta os resultados quando 50% dos nós da rede são egoístas. Percebemos que os resultados têm um leve acréscimo na taxa de acertos dos nós egoístas com mais de 99% de acerto mesmo quando $P_e = 75\%$. Esse resultado demonstra que mesmo aumentando a quantidade de nós egoístas, a taxa de acerto continua aumentando. Assim, concluímos que nosso modelo apresenta bons resultados principalmente quando há mais nós egoístas na rede. Embora nosso modelo não apresente grandes resultados para 10% dos nós da rede como egoísta, a separação da reputação nos agrupamentos ainda é satisfatória como verificado nas seções anteriores.

Figura 4.12: Taxa de verdadeiros positivos (TPR) para 25% dos nós da rede como egoístas.

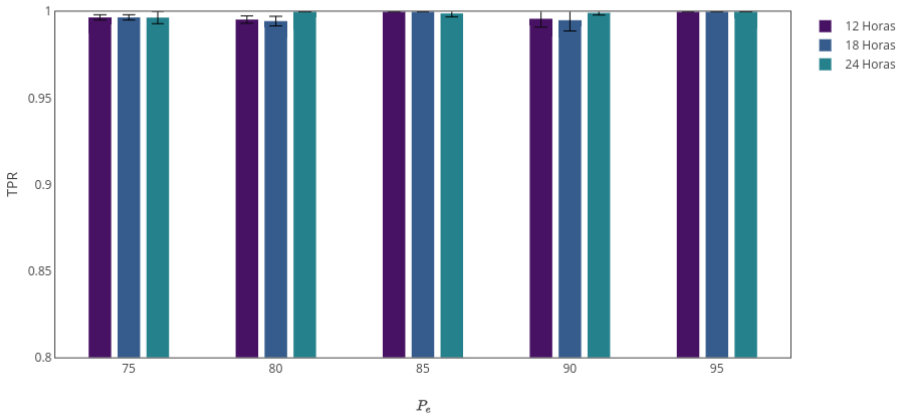
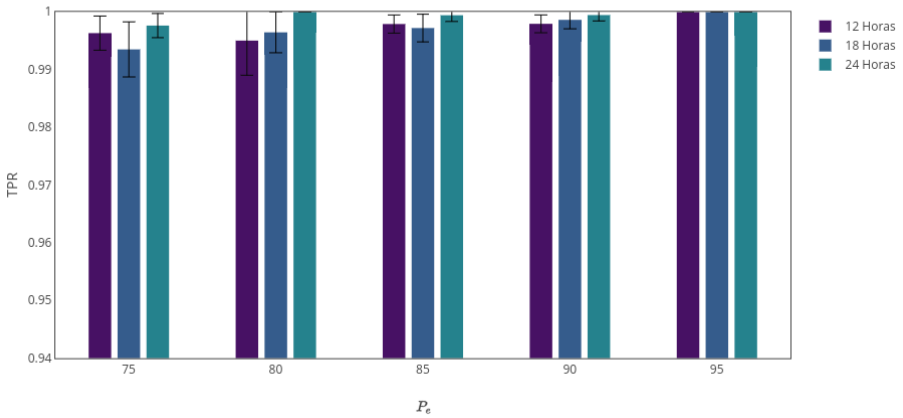


Figura 4.13: Taxa de verdadeiros positivos (TPR) para 50% dos nós da rede como egoístas.



Capítulo 5

Conclusões

O sucesso da comunicação em redes oportunistas depende principalmente da cooperação de todos os nós em prol do benefício coletivo. Contudo, alguns limites impostos nos recursos do dispositivo de cada usuário condicionam-os a se comportarem de modo a atender objetivos pessoais que, normalmente diferem dos objetivos gerais da rede. Modelos de reputação contribuem para a resolução desse problema, uma vez que permite aos membros da rede escolherem de modo eficaz estratégias para evitar nós egoístas ou incentivar o comportamento cooperativo. Para isso, realizam monitoramentos periódicos a fim de montar uma base de conhecimento a partir do histórico comportamental.

Neste trabalho, propomos uma arquitetura de detecção de nós egoístas utilizando um modelo de reputação para redes oportunistas, que possibilita aos nós da rede atestarem a integridade do comportamento dos nós. Esse mecanismo viabiliza atribuir mais pontuação para os nós que cooperam com a rede. Além disso, utilizamos uma metologia de agrupamento, também conhecida como *clustering* para realizar a classificação dos nós em egoístas ou não egoístas. Assim, não precisamos de informações suficientes sobre um nó até que a reputação desse nó ultrapasse certo limiar. Desse modo demonstramos que é preciso muito menos informações adquiridas para realizar a classificação dos nós.

Nossa arquitetura foi avaliada utilizando um *trace* de contato extraído de um experimento real com usuários e implementado no conhecido simulador The ONE. Através de um método de detecção analítico chamado *watchdog*, no qual a eficiência foi atribuída através do parâmetro $P_e \in \{75\%, 80\%, 85\%, 90\%, 95\%\}$ podemos ter mais controle sobre nosso experimento sem estar sujeito aos erros do *watchdog*. Cada detecção correta ou incorreta era utilizada para alimentar nosso modelo de reputação, que foi baseado nas funções logísticas, instrumentos comumente utilizados em diversos modelos de reputação na área de esportes, jogos e competições.

Nossos resultados demonstram que nossa proposta apresenta uma modelagem apropriada para indicar reputações para nós egoístas e não egoístas através dos estimadores de densidade *kernel*. Através destes estimadores percebemos que nosso modelo é bom para indicar nós com boas reputações, assim sendo um método apropriado para avaliar nós que ajudam com a cooperação. Embora a convergência da reputação dos nós egoístas ocorre de forma mais lenta, percebemos que o agrupamento é realizado com eficiência através do coeficiente de silhueta, uma métrica que avalia o grau de bom agrupamento.

Além disso realizamos um teste de classificação, no qual os nós classificam os nós em egoístas através dos valores de reputação. Nossos resultados demonstram acertos acima de 60% para 10% de nós egoístas e $P_e \in \{75\%, 80\%\}$ e acertos acima dos 90% quando $P_e \geq 85\%$. Quando aumentamos a quantidade de nós egoístas para 25% ou mais, nossos resultados obtêm taxas de acerto acima de 95%.

Com as situações apresentadas nos cenários de experimento, percebe-se claramente que nosso modelo tanto de reputação como a técnica utilizada para classificar os nós em egoístas através do agrupamento são ótimas estratégias que poderiam ser aplicadas no ambiente de redes oportunistas. Desde que o serviço de repasse na rede é amplamente relacionado a cooperação, é de suma importância a inclusão de mecanismos capazes de identificar comportamento egoísta.

Capítulo 6

Trabalhos Futuros

Nosso modelo aqui apresentado demonstra promissores resultados, no entanto, a metodologia aplicada na classificação utilizando agrupamento é nova na literatura de detecção de nós egoístas em redes oportunistas. Outras possíveis metodologias de aprendizagem de máquina não supervisionada podem obter mais êxito ou mesmo trazer futuros discernimentos em técnicas que podem ser aplicadas tais como lógica *fuzzy* ou o DBSCAN.

Além disso, a utilização de funções logísticas no âmbito da estratégia aplicada ao modelo de reputação indica comportamentos não esperados, como a baixa convergência da reputação dos nós egoístas mesmo quando o número de nós egoístas na rede é em torno de 50%. Assim, um estudo mais amplo sobre a inferência estatística envolvida nessa estratégia pode viabilizar futuras melhorias em nosso modelo.

A abordagem diferenciada de outros métodos de detecção presentes na literatura para o proposto aqui impediu algumas estratégias comparativas. Algumas equiparações entre estes modelos podem nos deu um conhecimento mais amplo no funcionamento e aplicabilidade de modelos de reputações em redes oportunistas.

Além disso, outras melhorias poderiam ter sido feitas no sistema proposto. Diversos parâmetros utilizados nos teste foram definidos de forma empírica baseados em estudos prévios, tais como a definição dos fatores utilizados no modelo de reputação, a probabilidade de eficiência no método de detecção e os parâmetros e atributos utilizados no agrupamento como o número de *clusters* $k = 2$. Estudos com $k = 3$ *clusters* podem ser aplicados futuramente para indicar um agrupamento no qual não é possível dizer se o nó é egoísta ou não egoísta, significando que pouca informação foi adquirida até o momento.

Outro estudo que podemos apontar como trabalho futuro é a verificação do comportamento da reputação de um nó quando os vizinhos mais próximos ou nós cuja

interação social é maior (amigos, proximidade social) possuem tendências para comportamento egoísta ou tendências para comportamento cooperativo. Este estudo pode servir para entender se a proximidade de um nó com outros nós egoístas pode influenciar a reputação dele com relação ao restante da rede.

Apêndice A

Lista de Publicações

A.1 Artigos Publicados ou Aceitos

1. SOUZA, C.; MOTA E. S.; CARVALHO, L. G.; SOARES, D. & NAVES, J. F. (2013). Uma Nova Política de Gerência de Buffer para Redes Tolerantes ao Atraso e Desconexões Baseada em Relações Sociais. In: *XXXI Simpósio Brasileiro de Telecomunicações, 2013.*

2. SOARES, D.; MOTA, E.; SOUZA, C.; MANZONI, P.; CANO, J. C. & CALAFATE, C. (2014). A statistical learning reputation system for opportunistic networks. In *Wireless Days (WD) 2014 IFIP*, (pp. 1-6). IEEE.

3. SOUZA, C.; MOTA, E.; SOARES, D.; MANZONI, P.; CANO, J. C. & CALAFATE, C. T. (2016). Improving delivery delay in social-based message forwarding in Delay Tolerant Networks. In: *Proceedings of the 2016 workshop on Fostering Latin-American Research in Data Communication Networks – LANCOMM '16*, pp. 52.

4. SOUZA, C.; MOTA, E. S.; CARVALHO, L. G. & SOARES, D.. (2017). Gerenciamento de Buffer Baseado em Egoísmo para Redes Tolerantes a Atrasos e Desconexões. In: *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2017).*

5. SOARES, D.; MATTHAUS, B.; MOTA, E. S. & CARVALHO, C. B. (2018). Avaliação Experimental da Eficácia de um Watchdog em Redes Oportunistas Móveis. In: *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2018).*

6. SOARES, D.; MATTHAUS, B.; MOTA, E. S. & CARVALHO, C. B. (2018). A Feasibility Study of Watchdogs on Opportunistic Mobile Networks. In: *2018 IEEE Symposium on Computers and Communications (ISCC) (ISCC 2018)*.

Referências Bibliográficas

Abdulla, M. & Simon, R. (2007). The Impact of the Mobility Model on Delay Tolerant Networking Performance Analysis. Em *40th Annual Simulation Symposium (ANSS'07)*, pp. 177--184. IEEE. ISSN 1080-241X.

Agrawal, D. (2005). A credit-payment scheme for packet forwarding fairness in mobile ad hoc networks. Em *IEEE International Conference on Communications, ICC 2005.*, volume 5, pp. 3005--3009. IEEE.

Azevedo, T. S.; Bezerra, R. L.; Campos, C. a. V. & de Moraes, L. F. M. (2009). An Analysis of Human Mobility Using Real Traces. *2009 IEEE Wireless Communications and Networking Conference*, pp. 1--6.

Balakrishnan, K. & Varshney, P. (2005). TWOAK: preventing selfishness in mobile ad hoc networks. Em *IEEE Wireless Communications and Networking Conference*, volume 4, pp. 2137--2142. IEEE. ISSN 1525-3511.

Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis. *SIAM Review*. ISSN 0036-1445.

Berkhin, P. (2006). A survey of clustering data mining techniques. Em *Grouping multidimensional data*, pp. 25--71. Springer.

Bessegato, L. F.; Atuncar, G. S. & Duczmal, L. H. (2006). Rotinas em r para técnicas de suavizaçao por núcleos estimadores. *IX Encontro de Modelagem Computacional, Anais*, pp. 1–10.

Bigwood, G. & Henderson, T. (2011). IRONMAN: Using Social Networks to Add Incentives and Reputation to Opportunistic Networks. Em *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pp. 65--72. IEEE.

Boldrini, C. & Passarella, A. (2010). HCMM: Modelling spatial and temporal properties of human mobility driven by usersâ social relationships. *Computer Communications*, 33(9):1056--1074. ISSN 01403664.

Braga, R. B. (2008). *MAPA: Mecanismo de Avaliação e Punição de nós egoístas em redes Ad hoc.* Tese de doutorado, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO.

Buchegger, S. & Le Boudec, J.-Y. (2002). Performance analysis of the CONFIDANT protocol. Em *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing - MobiHoc '02*, p. 226, New York, New York, USA. ACM Press.

Burt, R. S. (1992). *Structural Holes The Social Structure of Competition.* Harvard University Press.

Buttyan, L.; Holczer, T. & Schaffer, P. (2005). Spontaneous Cooperation in Multi-Domain Sensor Networks. *European Workshop on Security in Ad-hoc and Sensor Networks.*, pp. 42--53. ISSN 03029743.

Buttyán, L. & Hubaux, J.-P. (2003). Stimulating cooperation in self-organizing mobile ad hoc networks. *Mob. Netw. Appl.*, 8(5):579--592. ISSN 1383-469X.

Chaintreau, A.; Hui, P.; Crowcroft, J.; Diot, C.; Gass, R. & Scott, J. (2007). Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620. ISSN 1536-1233.

Charness, N.; Tuffiash, M.; Krampe, R.; Reingold, E. & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, 19(2):151--165.

Chen, H. & Chen, G. (2007). A resource-based reputation rating mechanism for peer-to-peer networks. Em *Sixth International Conference on Grid and Cooperative Computing (GCC 2007)*, pp. 535–541. ISSN 2160-4908.

Cho, E.; Myers, S. A. & Leskovec, J. (2011). Friendship and mobility. Em *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*, p. 1082, New York, New York, USA. ACM Press.

Cochran, W. G. (1954). Some methods for strengthening the common ² tests. *Biometrics*, 10(4):417–451. ISSN 0006341X, 15410420.

CRAWDAD (2014). Trace repository database at dartmouth. `http://crawdad.cs.dartmouth.edu/`.

Crowcroft, J. (2008). Predictability of human mobility and its impact on forwarding. Em *2008 Third International Conference on Communications and Networking in China*, pp. 543--547. IEEE.

Crowcroft, J.; Gibbens, R.; Kelly, F. & Östring, S. (2004). Modelling incentives for collaboration in mobile ad hoc networks. *Perform. Eval.*, 57(4):427--439. ISSN 0166-5316.

David B. Johnson, D. A. M. (1996). Dynamic Source Routing in Ad Hoc Wireless Networks. pp. 153--181.

Dawkins, R. (1976). The selfish gene.

de Souto, M.; Lorena, A.; Delbem, A. & de Carvalho, A. (2003). Técnicas de aprendizado de máquina para problemas de biologia molecular. *Sociedade Brasileira de Computaçao*.

Dini, G. & Lo Duca, A. (2010). A reputation-based approach to tolerate misbehaving carriers in Delay Tolerant Networks. Em *The IEEE symposium on Computers and Communications*, pp. 772--777. IEEE. ISSN 1530-1346.

Eagle, N. & (Sandy) Pentland, A. (2005). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255--268. ISSN 1617-4909.

Fall, K. (2003). A delay-tolerant network architecture for challenged internets. *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications - SIGCOMM '03*, p. 27.

Furletti, B.; Gabrielli, L.; Renso, C. & Rinzivillo, S. (2013). Analysis of GSM calls data for understanding user mobility behavior. *2013 IEEE International Conference on Big Data*, pp. 550--555.

Gasch, A. P. & Eisen, M. B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome biology*, 3(11):research0059--1.

Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J. P.; Coller, H.; Loh, M. L.; Downing, J. R.; Caligiuri, M. A. & Bloomfield, C. D. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286:531--537.

Haggle (2014). Haggle project. `https://code.google.com/p/haggle/`.

Hair, J.; Anderson, R. & Tatham, R. (2005). *Analise Multivariada de Dados*. Bookman. ISBN 9788536304823.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P. & Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10--18. ISSN 1931-0145.

Hartigan, J. A. & Wong, M. A. (1979). A k-means clustering algorithm. *JSTOR: Applied Statistics*, 28(1):100--108.

He, Q.; Wu, D. & Khosla, P. (2004). Sori: a secure and objective reputation-based incentive scheme for ad-hoc networks. Em *2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*, volume 2, pp. 825–830 Vol.2. ISSN 1525-3511.

Hernández-Orallo, E.; Olmos, M. D. S.; Cano, J.-C.; Calafate, C. T. & Manzoni, P. (2013). A Fast Model for Evaluating the Detection of Selfish Nodes Using a Collaborative Approach in MANETs. *Wireless Personal Communications*, 74(3):1099--1116. ISSN 0929-6212.

Hernández-Orallo, E.; Olmos, M. D. S.; Cano, J. C.; Calafate, C. T. & Manzoni, P. (2015). CoCoWa: A collaborative contact-based watchdog for detecting selfish nodes. *IEEE Transactions on Mobile Computing*. ISSN 15361233.

Holliday, J. D.; Rodgers, S. L.; Willett, P.; Chen, M.-Y.; Mahfouf, M.; Lawson, K. & Mullier, G. (2004). Clustering files of chemical structures using the fuzzy k-means clustering method. *Journal of chemical information and computer sciences*, 44(3):894--902.

Hui, P. & Crowcroft, J. (2007). How Small Labels Create Big Improvements. Em *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07)*, pp. 65--70. IEEE.

Hui, P.; Crowcroft, J. & Yoneki, E. (2011). BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks. *IEEE Transactions on Mobile Computing*, 10(11):1576--1589. ISSN 1536-1233.

Ii, F.; Wang, M. & Abdeldjalil, T. (2013). SEBAR : Social Energy BAsed Routing Scheme for Mobile Social Delay Tolerant Networks.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651--666.

Jain, A. K. & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA. ISBN 0-13-022278-X.

Jenks, G. F. (1967). The data model concept in statistical mapping. *International yearbook of cartography*, 7:186--190.

Karagiannis, T.; Le Boudec, J.-Y. & Vojnovic, M. (2010). Power Law and Exponential Decay of Intercontact Times between Mobile Devices. *IEEE Transactions on Mobile Computing*, 9(10):1377--1390. ISSN 1536-1233.

Karaliopoulos, M. (2009). Assessing the vulnerability of DTN data relaying schemes to node selfishness. *IEEE Communications Letters*, 13(12):923--925. ISSN 1089-7798.

Kato, N. & Nemoto, Y. (1996). Large scale hand-written character recognition system using subspace method. Em *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, volume 1, pp. 432–437 vol.1. ISSN 1062-922X.

Kaufman, L. & Rousseeuw, P. J. (2008). *Partitioning Around Medoids (Program PAM)*, pp. 68--125. John Wiley Sons, Inc.

Kejun Liu; Jing Deng; Varshney, P. & Balakrishnan, K. (2007). An Acknowledgment-Based Approach for the Detection of Routing Misbehavior in MANETs. *IEEE Transactions on Mobile Computing*, 6(5):536--550. ISSN 1536-1233.

Keränen, A.; Ott, J. & Kärkkäinen, T. (2009). The one simulator for dtn protocol evaluation. Em *Proceedings of the 2Nd International Conference on Simulation Tools and Techniques*, Simutools '09, pp. 55:1--55:10, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Kim, M.; Kotz, D. & Kim, S. (2006). Extracting a Mobility Model from Real User Traces. Em *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pp. 1--13. IEEE. ISSN 0743-166X.

Krifa, A. (2012). *Towards Efficient Content Dissemination Over Disruption Tolerant Networks*.

Le, V.; Scholten, J.; Havinga, P. & Ngo, H. (2014). Location-based Data Dissemination with Human Mobility Using Online Density Estimation. *2014 IEEE Consumer Communications & Networks Conference*.

Lee, K.; Hong, S.; Kim, S.; Rhee, I. & Chong, S. (2008). Demystifying levy walk patterns in human walks. *Techical Report, NCSU*.

Lee, K.; Hong, S.; Kim, S. J.; Rhee, I. & Chong, S. (2009). SLAW: A New Mobility Model for Human Walks. Em *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pp. 855--863. IEEE. ISSN 0743-166X.

Li, N. & Das, S. K. (2010). RADON: reputation-assisted data forwarding in opportunistic networks. Em *Proceedings of the Second International Workshop on Mobile Opportunistic Networking - MobiOpp '10*, p. 8, New York, New York, USA. ACM Press.

Li, Q.; Zhu, S. & Cao, G. (2010). Routing in Socially Selfish Delay Tolerant Networks. Em *2010 Proceedings IEEE INFOCOM*, pp. 1--9. IEEE. ISSN 0743-166X.

Li, Y.; Su, G. & Wang, Z. (2012). Evaluating the effects of node cooperation on DTN routing. *AEU - International Journal of Electronics and Communications*, 66(1):62--67. ISSN 14348411.

Li, Y.; Su, G.; Wu, D. O.; Jin, D.; Su, L. & Zeng, L. (2011). The Impact of Node Selfishness on Multicasting in Delay Tolerant Networks. *IEEE Transactions on Vehicular Technology*, 60(5):2224--2238. ISSN 0018-9545.

Liu, M.; Yang, Y. & Qin, Z. (2011). A survey of routing protocols and simulations in delay-tolerant networks. pp. 243--253.

Lu, R.; Lin, X.; Zhu, H.; Shen, X. & Preiss, B. (2010). Pi: A practical incentive protocol for delay tolerant networks. *IEEE Transactions on Wireless Communications*, 9(4):1483--1493. ISSN 1536-1276.

Luo, P.; Huang, H.; Shu, W.; Li, M. & Wu, M.-Y. (2008). NET 07-2 - Performance Evaluation of Vehicular DTN Routing under Realistic Mobility Models. Em *2008 IEEE Wireless Communications and Networking Conference*, pp. 2206--2211. IEEE. ISSN 1525-3511.

Marti, S.; Giuli, T. J.; Lai, K. & Baker, M. (2000). Mitigating routing misbehavior in mobile ad hoc networks. Em *Proceedings of the 6th annual international conference*

on Mobile computing and networking - MobiCom '00, pp. 255--265, New York, New York, USA. ACM Press.

McAuley, J. & Leskovec, J. (2012). Discovering Social Circles in Ego Networks. p. 30.

McNett, M. & Voelker, G. M. (2005). Access and mobility of wireless PDA users. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(2):40. ISSN 15591662.

Michiardi, P. & Molva, R. (2002). Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. pp. 107--121.

Miranda, H. & Rodrigues, L. (2003). Friends and foes: preventing selfishness in open mobile ad hoc networks. Em *Distributed Computing Systems Workshops, 2003. Proceedings. 23rd International Conference on*, pp. 440–445.

Mouly, M. & Pautet, Marie-Bernadette/Foreword By-Haug, T. (1992). The GSM System for Mobile Communications.

Neto, J. B. P. (2011). Um Modelo para Previsão do Volume de Contato em Redes Tolerantes a Atrasos e Desconexões: Uma Abordagem Quantitativa. *Universidade Federal do Amazonas*.

Oliveira, C. T.; Moreira, M. D. D.; Rubistein, M. G.; Costa, L. H. M. K. & Duarte, O. C. M. B. (2007). Redes Tolerantes a Atrasos e Desconexões. Em *Short courses in 25th Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*, capítulo 5, pp. 203--256.

Panagakis, A.; Vaios, A. & loannis Stavrakakis (2007). On the Effects of Cooperation in DTNs. Em *2007 2nd International Conference on Communication Systems Software and Middleware*, pp. 1--6. IEEE.

Pujol, J. M.; Lopez Toledo, A. & Rodriguez, P. (2009). Fair Routing in Delay Tolerant Networks. Em *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pp. 837--845. IEEE. ISSN 0743-166X.

Rollernet (2014). `http://www-rp.lip6.fr/rollernet/en/index.html`.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(Supplement C):53--65. ISSN 0377-0427.

Scott, D. W. (2008). Kernel density estimators. *Multivariate Density Estimation: Theory, Practice, and Visualization*, pp. 125--193.

Scott, K., Burleigh, S. (2007). Bundle Protocol Specification. RFC5050.

Silva, B. M. C.; Rodrigues, J. J. P. C.; Proenç, M. L.; Kumar, N.; Proença, M. L.; Han, G.; Silva, B. M. C.; Rodrigues, J. J. P. C. & Proença, M. L. (2016). MobiCoop: An Incentive-Based Cooperation Solution for Mobile Applications. *ACM Trans. Multimedia Comput. Commun. Appl. Article ACM Trans. Multimedia Comput. Commun. Appl*, 12(49).

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Slocum, T. A. & Egbert, S. L. (1993). Knowledge acquisition from choropleth maps. *Cartography and Geographic Information Systems*, 20(2):83--95.

Soares, D.; Mota, E.; Souza, C.; Manzoni, P.; Cano, J. C. & Calafate, C. (2015). A statistical learning reputation system for opportunistic networks. Em *IFIP Wireless Days*. ISSN 2156972X.

Socievole, A.; De Rango, F. & Marano, S. (2013). Face-to-face with facebook friends: Using online friendlists for routing in opportunistic networks. Em *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 2989--2994. IEEE. ISSN 2166-9570.

Souza, C. B.; Mota, E. & Soares, D. (2017). Gerenciamento de Buffer Baseado em Egoísmo para Redes Tolerantes a Atrasos e Desconexões. *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2017)*, pp. 301--314.

Spyropoulos, T.; Psounis, K. & Raghavendra, C. S. (2006). Performance analysis of mobility-assisted routing. Em *Proceedings of the seventh ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '06*, p. 49, New York, New York, USA. ACM Press.

Srinivasan, V.; Nuggehalli, P.; Chiasserini, C. & Rao, R. (2003). Cooperation in wireless ad hoc networks. Em *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pp. 808–817 vol.2. ISSN 0743-166X.

Toh, C.; Kim, D.; Oh, S. & Yoo, H. (2010). The controversy of Selfish nodes in ad hoc networks. 2:1087--1092. ISSN 1738-9445.

Tournoux, P.-U.; Leguay, J.; Benbadis, F.; Conan, V.; Dias de Amorim, M. & Whitbeck, J. (2009). The Accordion Phenomenon: Analysis, Characterization, and Impact on DTN Routing. Em *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pp. 1116--1124. IEEE. ISSN 0743-166X.

USC (2014). Repositorio de traces coletados na ncsu. `http://www.lib.ncsu.edu/dli/projects/mobilib`.

Vakali, A. & Pallis, G. (2007). *Web data management practices : emerging techniques and technologies*. Idea Group Pub. ISBN 9781599042282.

Wang, Y.; Dou, Q.; Peng, W. & Gong, Z. H. (2013). DTN Routing Performance Evaluation under Random Waypoint with Base Point Mobility Model. *Applied Mechanics and Materials*, 411-414:676--679. ISSN 1662-7482.

Wu, G.; Wang, J.; Yao, L. & Lin, C. (2013). A Secure Social-Aware Incentive Scheme for Delay Tolerant Networks. Em *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 813--820. IEEE.

Wu, Y.; Deng, S. & Huang, H. (2012). On Modeling The Impact of Selfish Behaviors on Limited Epidemic Routing in Delay Tolerant Networks. *Wireless Personal Communications*. ISSN 0929-6212.

Zhong, S.; Chen, J. & Yang, Y. (2003). Sprite: a simple, cheat-proof, credit-based system for mobile ad-hoc networks. Em *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pp. 1987–1997 vol.3. ISSN 0743-166X.

Zhu, K.; LI, W. & Fu, X. (2014). SMART: A Social and Mobile Aware Routing Strategy for Disruption Tolerant Networks. *IEEE Transactions on Vehicular Technology*, PP(99):1--1. ISSN 0018-9545.