



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
ESCOLA DE INFORMÁTICA APLICADA

Uma análise da conectividade de dispositivos sem fio através de dados reais de mobilidade humana

Yuri Menezes e Guilherme Oliveira

Orientador: Carlos Alberto

Rio de Janeiro

Novembro de 2013

Agradecimentos

A Deus, Rei sobre tudo, a quem tudo devemos, pela benção das amizades criadas e do conhecimento adquirido;

A Elenilson Gomes, que nos ajudou no tratamento e conversão dos dados coletados;

A Alexandre Silva e Edmilson Rocha, que nos ajudaram na análise estatística dos resultados obtidos;

Às nossas famílias, que não nos deixaram perder o foco mediante as dificuldades encontradas;

Ao professor Carlos Alberto Campos, que nos proporcionou a oportunidade de trabalhar nesse projeto e nos orientou ao longo do caminho.

Sumário

| | |
|---|----|
| 1. Introdução | 1 |
| 2. Fundamentação Teórica..... | 4 |
| 2.1 Estrutura e Arquitetura DTN | 4 |
| 2.2 Roteamento..... | 7 |
| 3. Trabalhos Relacionados..... | 12 |
| 4. Investigação sobre a conectividade dos nós a partir do uso de traces de contatos sociais | 15 |
| 4.1 Escolha dos dados..... | 15 |
| 4.2 Os Cenários Analisados..... | 16 |
| 5. Metodologia utilizada para análise da conectividade | 18 |
| 5.1 Rotina de transformação | 18 |
| 5.2 Geração das medidas | 20 |
| 6. Resultados obtidos..... | 25 |
| 6.1 Rollernet | 26 |
| 6.2 Infocom5..... | 34 |
| 6.3 Infocom6..... | 41 |
| 6.4 Análise das Métricas no tempo..... | 49 |
| 7. Conclusão | 54 |
| 7.1 Considerações Finais | 54 |
| 7.2 Dificuldades encontradas ao longo deste trabalho | 57 |
| 8. Referências Bibliográficas..... | 58 |

Sumário de Imagens

| | |
|--|----|
| Figura 1: Arquitetura da rede DTN (fonte: [1])..... | 5 |
| Figura 2: Ideologia de propagação de informação do Bubble Rap (fonte: [15])..... | 9 |
| Figura 3: Exemplo de matriz de adjacência | 19 |
| Figura 4: Exemplo do cálculo do coeficiente de clusterização (fonte: [3])..... | 23 |
| Figura 5: Análise da Densidade no cenário Rollernet | 27 |
| Figura 6: Análise do Grau Médio no cenário Rollernet | 28 |
| Figura 7: Análise de Variação do Grau dos nós no cenário Rollernet | 30 |
| Figura 8: Análise do Coeficiente de Agrupamento no cenário Rollernet..... | 31 |
| Figura 9: Análise da Frequência de Duração dos Contatos no cenário Rollernet | 33 |
| Figura 10: Análise da Densidade no cenário Infocom5 | 35 |
| Figura 11: Análise do Grau Médio no cenário Infocom5..... | 36 |
| Figura 12: Análise de Variação do Grau dos nós no cenário Infocom5..... | 38 |
| Figura 13: Análise do Coeficiente de Agrupamento no cenário Infocom5 | 39 |
| Figura 14: Análise da Frequência de Duração dos Contatos no cenário Infocom5 | 40 |
| Figura 15: Análise da Densidade no cenário Infocom6 | 42 |
| Figura 16: Análise do Grau Médio no cenário Infocom6..... | 44 |
| Figura 17: Análise de Variação do Grau dos nós no cenário Infocom6..... | 46 |
| Figura 18: Análise do Coeficiente de Agrupamento no cenário Infocom6..... | 47 |
| Figura 19: Análise da Frequência de Duração dos Contatos no cenário Infocom6 | 48 |
| Figura 20: Análise dos resultados no tempo das métricas no cenário Rollernet | 50 |
| Figura 21: Análise dos resultados no tempo das métricas no cenário Infocom5 | 51 |
| Figura 22: Análise dos resultados no tempo das métricas no cenário Infocom6 | 52 |

Sumário de Tabelas

| | |
|--|----|
| Tabela 1: Dados dos cenários | 16 |
| Tabela 2: Resultados do MSE das métricas do cenário Rollernet | 26 |
| Tabela 3: Resultados do MSE da variação do grau dos nós no cenário Rollernet | 30 |
| Tabela 4: Resultados do MSE das métricas do cenário Infocom5 | 34 |
| Tabela 5: Resultados do MSE da variação do grau dos nós no cenário Infocom5..... | 37 |
| Tabela 6: Resultados do MSE das métricas do cenário Infocom6 | 42 |
| Tabela 7: Resultados do MSE da variação do grau dos nós no cenário Infocom6..... | 45 |
| Tabela 8: Resultados da Análise do comportamento das curvas ECDF..... | 55 |

Notações

D : densidade do grafo da rede (pág. 20);

A : conjunto das arestas dos grafos da rede (pág. 20);

C : número total de arestas contidas em A (pág. 20);

N : conjunto dos nós dos grafos da rede (pág. 20);

X : número total de nós contidos em N (pág. 20);

$G(A,N)$: grafo da rede formado pelo conjunto A de arestas e o conjunto N de nós (pág. 20);

t : o instante da simulação de rede analisado (pág. 20);

g : grau do nó, representando o número de arestas que saem dele (pág. 21);

gm : o grau médio do grafo da rede (pág. 21);

i : um nó específico analisado (pág. 21);

C_i : o coeficiente de agrupamento do nó i (pág. 22);

k_i : o número de nós vizinhos ao nó i (pág. 22).

1. Introdução

Atualmente, a área de redes sem fio está sendo bastante estudada, visto que, em quase todo lugar existe sua utilização. Seu uso se torna mais facilitado por não necessitar de instalação e utilização de cabos, apenas a necessidade de um ponto de acesso remoto.

As redes sem fio são divididas em duas categorias: com infra-estrutura (WLAN, WMAN, etc.) e sem infra-estrutura (redes celulares, redes veiculares, redes Ad-Hoc, redes tolerantes a atraso, etc.). Um desses tipos de redes sem infra-estrutura será o objeto de estudo deste trabalho: as redes tolerantes a atraso e interrupções (DTN).

Segundo [1], as redes DTN emergiram da escuridão para um importante tópico de pesquisa tanto de arquitetos de redes quanto de desenvolvedores de aplicações, graças aos trabalhos realizados com MANETs (*Mobile Ad-Hoc Networks*). Esse tipo de rede é necessário para situações onde não é possível uma conexão contínua fim-a-fim, visto que não é necessário um pré-estabelecimento de conexão para seu funcionamento.

Uma vantagem das redes DTN é, justamente, a falta de necessidade de um pré-estabelecimento de uma conexão entre o emissor e receptor da mensagem a ser enviada, como, por exemplo, ocorre em redes sob protocolo TCP/IP. Uma desvantagem das redes DTN é a falta de garantia de que o conteúdo a ser enviado chegará ao seu destino. Nem ao menos se pode estimar um tempo de viagem da informação, pois não se sabe que caminho ela irá seguir até chegar a seu destinatário. Além disso, há um problema de segurança da informação, pois os métodos de criptografia (codificação da informação) tradicionais utilizados, por exemplo, em redes sob o protocolo TCP-IP, não são aplicados nas redes DTN.

O objetivo deste trabalho é a análise do comportamento das conexões sociais, as quais trataremos por ‘contatos’, mediante uso de grafos de conectividade, que é a estrutura que ilustra a localização e o contato entre os nós, visando compreendê-lo, obtendo informação sobre suas características, e, assim, permitir que futuras aplicações a serem desenvolvidas, baseadas em redes DTN, sejam mais eficientes. Serão analisados cenários reais de mobilidade humana, através da geração de grafos de conectividade e implementação de métricas que ilustrarão o comportamento dos contatos nos cenários analisados, sendo o instrumento principal de análise para que consigamos chegar a importantes conclusões.

Este trabalho é dividido em sete capítulos, que possuem seções e subseções para explicar, com maior nível de detalhamento, aspectos relativos ao assunto abordado pelo capítulo em questão. O Capítulo 1 está sendo esta introdução.

O Capítulo 2 terá como assunto a fundamentação teórica, ou seja, os conceitos básicos que serão utilizados neste trabalho para chegar ao seu objetivo final: a análise de dados de cenários de mobilidade humana, baseada em implementação de métricas. Esses cenários são baseados na arquitetura das redes DTN, assunto principal deste capítulo. Esse capítulo é dividido em duas seções, onde a primeira foca na estrutura e conceitos básicos desse tipo de rede, e a segunda, nas formas de transmissão de informação, seus protocolos de roteamento, tanto os mais usados quanto os menos conhecidos.

O assunto a ser abordado no Capítulo 3 será uma listagem de trabalhos anteriormente realizados que possuem algum assunto em comum com o abordado neste trabalho. São comentados os principais conceitos de cada um, sendo os que mais se assemelham ao tratado nesta monografia e que, conseqüentemente, serviram de embasamento teórico e prático.

O Capítulo 4 contém o primeiro passo da parte prática deste trabalho: a escolha dos cenários a serem analisados. Ele também foi dividido em duas seções: a primeira enfocando o repositório de onde os cenários foram coletados, em que extensões de arquivos esses cenários são disponibilizados e etc; a segunda explicando cada cenário analisado, tirando os principais dados que serão utilizados na análise final dos resultados.

O Capítulo 5 contém o detalhamento das práticas utilizadas sobre os cenários, desde o tratamento inicial até a análise final dos dados dos cenários. Ele é dividido em três seções. A primeira seção dá enfoque à rotina de transformação implementada para converter os dados coletados do repositório para um determinado padrão. Padrão esse que foi base para a implementação das métricas, geradoras dos resultados que serão o objeto da análise final. A segunda seção mostra quais métricas foram escolhidas para a análise, suas definições e fórmulas e sua importância dentro do cenário.

O Capítulo 6 mostra todo o conteúdo sobre a análise, desde a explicação de como ela foi feita até seus resultados e conclusões tiradas a partir deles. A seção foi dividida em quatro subseções, as três primeiras descrevendo os resultados de cada cenário e a última descrevendo uma análise dos resultados no tempo.

O Capítulo 7 será o da conclusão, que terá um resumo do que foi feito no trabalho, com as conclusões que foram tiradas mediante os resultados da análise e as dificuldades encontradas durante o trabalho.

E, por fim, o Capítulo 8 será destinado às referências bibliográficas. Ele conterá todas as teses, artigos e outros trabalhos utilizados como embasamento teórico para definições dentro desta monografia.

2. Fundamentação Teórica

Nesta seção, serão mostradas definições importantes sobre o conceito das redes DTN, desde sua estrutura básica (arquitetura) até seus mecanismos internos (segurança e roteamento, por exemplo).

2.1 Estrutura e Arquitetura DTN

As redes DTN (*Delay-tolerant networking*) são redes tolerantes a atraso e desconexão. Sua arquitetura foi projetada para acomodar não apenas a interrupção de conexão de rede, mas também para fornecer uma estrutura para lidar com os tipos de heterogeneidade encontrados nos *gateways* de rede de sensores.

O DTN pode usar uma grande variedade de protocolos de entrega diferentes durante a transmissão dos dados entre os seus nós intermediários ou finais, incluindo o TCP / IP, mesmo sendo um caso raro, dado que as ideias gerais das redes TCP-IP e redes DTN são totalmente opostas.

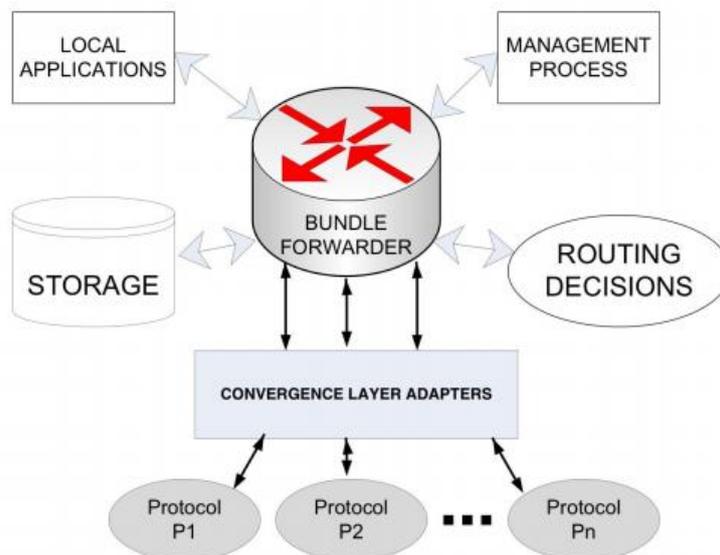


Figura 1: Arquitetura da rede DTN (fonte: [1])

Como mostra a Figura 1, devido ao fato de cada um destes protocolos fornecer uma semântica um pouco diferente, uma coleção de adaptadores de protocolos específicos de camada de convergência (*Convergence Layer Adapters*) é utilizada. Ela proporciona as funções necessárias para transportar unidades de dados de protocolo DTN (chamado *bundles*) em cada um dos protocolos correspondentes. A utilização do “*bundle*” para reunir toda a informação necessária para concluir uma transação (por exemplo, opções de protocolo e dados de autenticação), possibilita a redução do número de trocas de protocolos (intercâmbios).

A arquitetura DTN também provê um processo de gerenciamento (*Management Process*) que é responsável, por exemplo, pela definição da rota (*Routing Decisions*) que cada informação irá utilizar para chegar a seu destinatário. Os agentes remetentes e receptores de informação são aplicações locais (*Local Applications*). Cada nó possui um espaço de armazenamento (*Storage*) para manter informação enviada a ele, ou que está sob sua custódia.

Para facilitar o transporte de *bundles* com uma quantidade grande de dados, é utilizado o processo de fragmentação, ou seja, cada *bundle* é dividido em subpartes e cada parte então, mandada separadamente para seu destino. A fragmentação ajuda a evitar congestionamento de rede. Um cabeçalho especial é usado para relacionar um fragmento a seu *bundle* original. Para o provimento de segurança de informação, um processo de criptografia também pode ser usado junto com o de fragmentação.

Existem dois blocos para definição de integridade de dados: um para integridade fim-a-fim (sem o uso de entidades intermediárias) e outro para integridade '*hop-by-hop*' (quando o *bundle* passa por nós intermediários antes de alcançar seu destino). Nós intermediários relevantes, chamados fonte ou destino seguro (que podem diferir da fonte e destino do *bundle*), têm permissão em DTN para aplicar ou checar a validade das credenciais criptográficas do *bundle*.

Em geral, para garantir segurança à informação trafegada, as redes utilizam alguns processos, sendo um deles, a criptografia. Entretanto, o protocolo de segurança de *bundles* não provê qualquer forma de gerenciamento das chaves criptográficas necessárias. Um estudo de um modelo de autorização de tráfego - análogo à solução dada ao problema de autenticação, autorização e contabilidade na Internet – ainda se faz necessário em DTN.

Nas redes DTN, existe um procedimento chamado de transferência de custódia. Trata-se de um serviço opcional que pode ser oferecido a um *bundle* durante sua entrega em uma DTN. Consiste em delegar a uma entidade responsável identificada a função de manter o *bundle* em segurança até que outra entidade o receba. Os nós DTN possuem a opção de executar ou não esse processo de transferência de custódia. Um nó pode negá-lo devido à sua implementação, políticas, falta de espaço, entre outras razões. Um problema desse procedimento é a impossibilidade de uma rede DTN operar com transferência de custódia em todos os seus nós, o que seria o ideal em uma rede estável e com recursos de armazenamento suficientes.

Quando os recursos de armazenamento se tornam escassos devido à presença excessiva de dados ou fragmentos de *bundles*, diz-se que ocorre um congestionamento. Ele interfere diretamente na eficiência da rede, tornando-a cada vez mais lenta, dependendo do tamanho do congestionamento. Quanto maior o congestionamento, mais lento fica. Uma maneira para evitar o congestionamento é a fragmentação, mencionada anteriormente nesta monografia.

As redes DTN utilizam o processo de criptografia a fim de prover segurança no tráfego de seus dados nos *bundles*. Para criptografar, o *Protocolo Bundle* utiliza a técnica SDNV (*Self-Delimiting Numeric Values*) que consiste em codificar valores numéricos positivos que podem se estender em grande variedade. Esses valores atuam similarmente a senhas. Apesar de prover serviços de criptografia, o protocolo de segurança de *bundles* não provê qualquer forma de gerenciamento das chaves criptográficas necessárias, como, por exemplo, chaves simétricas e assimétricas (pública e privada), que são utilizadas nas redes TCP-IP, por exemplo.

2.2 Roteamento

A área de estudo sobre roteamento é uma das mais importantes no conceito de redes de computadores, pois ela é o fator crucial para apontar se um protocolo de rede é bom ou ruim, ou seja, caso o roteamento seja excelente, muito possivelmente essa rede também será.

Em redes DTN, existem muitos métodos para realizar o roteamento, que podem envolver a criação e destruição de cópias de um *bundle*, sua fragmentação, granularidade na tomada de decisões, reserva de recursos, etc.

O roteamento possui uma tendência à literatura *Mobile Ad-hoc Network* (MANET), que foca em roteamento em redes móveis *Ad-Hoc* relativamente densas, onde há possibilidade de conexão ponto-a-ponto entre qualquer par de nós. Neste trabalho, trataremos como ‘nós’ os emissores, receptores ou entidades entre eles pelas quais o *bundle* passou.

Essas entidades intermediárias podem existir pelo fato de que, como em redes DTN não há a obrigatoriedade de uma pré-conexão, pode não haver uma transmissão direta, simplesmente por não se saber onde está o receptor. Então, em cada nó, é feito um novo cálculo de rota, descobrindo que nó mais próximo servirá para passar o *bundle* até que ele chegue ao seu destino. Existem casos em que o receptor pode ser encontrado sem precisar de nós intermediários, mas é um caso muito raro. Para que um nó saiba da existência de seus vizinhos, periodicamente, cada um deles emite uma mensagem, atuando como um sinal de sua existência.

No futuro, nós DTN deverão suportar diferentes estratégias e protocolos de roteamento a fim de operar eficientemente nos diversos ambientes nos quais se encontrarão, visando atender todas as necessidades dos seus usuários da mesma maneira, ou melhor, que as redes sob o protocolo TCP-IP atendem aos seus.

A seguir, serão mencionados alguns dos algoritmos de roteamento mais utilizados em redes DTN:

- *Bubble Rap*: Este algoritmo [15], proposto por estudantes na Universidade de Cambridge, leva em consideração estruturas e centralidades na hora de escolher por qual caminho uma informação será propagada.

A Figura 2, a seguir, nos mostra essas estruturas, chamadas de comunidades, que são um grupo de nós que frequentemente trocam informação entre si. Em cada comunidade, existe aquele nó que é mais propenso a participar de uma troca de informação, ou seja, que possui a maior centralidade. Neste algoritmo, a estratégia para tomar a decisão da rota que a informação irá tomar é: como primeira parte da estratégia, enviar a mensagem para um nó mais popular (maior centralidade) que o nó emissor e, como segunda parte, reconhecer os membros da comunidade destino e utilizados como propagadores da informação. Então, esses nós repassam a informação para nós com maior nível de centralidade dentro das sub-comunidades, que são grupos de nós dentro das comunidades. A partir daí, começa uma nova propagação da informação.

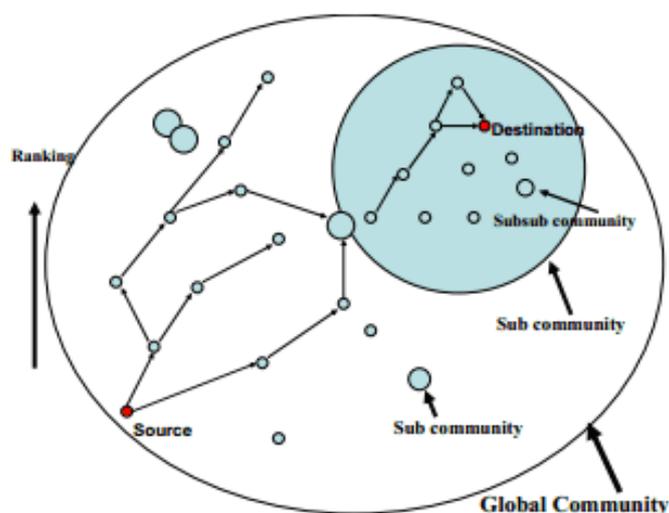


Figura 2: Ideologia de propagação de informação do Bubble Rap (fonte: [15])

- *Epidemic*: Este algoritmo [16] tem, como princípio de estratégia de decisão de rota, a ideia mais básica para a disseminação da informação. Consiste em enviar cópias da informação para todos os nós com os quais o emissor possui contato, e assim, contar com a retransmissão executada por esses nós para que a mensagem se propague pela rede.

Essa metodologia de envio possui um problema. Devido ao fato de a mesma informação ser disseminada por todos os nós, gera-se um acúmulo muito grande de mensagens com informação repetida, com conseqüente necessidade de muitos *buffers* e risco de perda de dados. No âmbito do uso de simuladores de rede, o uso desse protocolo aumenta, consideravelmente, o tempo de simulação.

- *Spray and Wait*: Este algoritmo [17] tem, como princípio de estratégia de decisão de rota, a ação de disseminar (*spray*) um determinado número de cópias da informação e esperar (*wait*) que algum dos nós receba essa informação.

No âmbito da teoria e de simulações, esse algoritmo supera grande parte dos outros existentes com relação à média de atraso na entrega da informação e número de transmissões por informação entregue. Seu desempenho geral é perto do ideal. Além disso, é um algoritmo escalável, que mantém bom desempenho em uma variedade de cenários, ao contrário de outros algoritmos, gerando também um tempo de simulação menor.

Outros exemplos, menos utilizados, de algoritmos de roteamento em redes DTN são:

- *Socio-Aware Overlay*: Proposto por Yoneki [2], é uma técnica que leva em consideração redes com infraestrutura. Consiste em criar uma sobreposição de uma rede oportunística de comunicação *publish/subscribe*. Essa sobreposição é composta por nós com um alto valor de centralidade que possuem a melhor visibilidade em uma comunidade (os *hubs* de cada uma). Comunidade se refere a um grupo de nós que são unidos baseados nas fortes relações sociais entre os mesmos, ou seja, nós que frequentemente trocam *bundles* muito provavelmente pertencem à mesma comunidade. Nós de uma comunidade podem se comunicar com outros de outra comunidade (comunidades familiarizadas).

- *Wireless Ad-Hoc Podcasting*: Apresentado por Lenders [9] [10], esse protocolo tem como propósito a entrega de conteúdo, por meio sem-fio, entre os nós móveis. Assumindo uma rede sem infraestrutura, o serviço de *podcasting* sem fio permite a distribuição de conteúdos através de contatos oportunistas sempre que os dispositivos estão em uma faixa de comunicação sem fio. Do ponto de vista de organização do conteúdo, o *Ad Hoc Podcasting* emprega uma abordagem *publish/subscribe*. Assim, ele organiza o conteúdo em canais, que permite aos usuários assinar e receber automaticamente as atualizações para o conteúdo em que estão interessados.
- DPSP: Proposto por Greifenberg e Kutscher [11], esse protocolo é uma infraestrutura probabilística de distribuição *multicast* baseado em *publish/subscribe* para redes DTN. O protocolo utiliza um modelo de entrega de distribuição assíncrono baseado em empurrão. A ideia é que cada nó na rede replique *bundles* para seus vizinhos de forma que esse *bundle* seja entregue por vários saltos do tipo *store-carry-and-forward*. O DPSP tem uma organização de conteúdo com base em um sistema de subscrição de canais, onde os usuários se inscrevem em canais e remetentes publicam conteúdo neles.
- *ContentPlace*: Proposto por Boldrini [12], esse protocolo lida com a disseminação em redes oportunistas com recursos limitados, disponibilizando o conteúdo onde usuários interessados possam estar, poupando assim os recursos disponíveis. Aprendendo sobre as relações sociais dos usuários, é possível determinar o melhor local para disponibilizar o conteúdo. Assumindo uma rede sem infraestrutura, em *ContentPlace* um nó decide que informação deve ser replicada localmente a cada vez que encontra outro nó, e existe uma política de replicação. Estando ao alcance um do outro, dois nós devem primeiro descobrir um ao outro; a forma não é especificada, mas é provável que usem *broadcasts* por *Wi-Fi* ou *Bluetooth*, já que os nós são nada mais que dispositivos móveis, e para identificação de conteúdo, o nó deverá avisar em que grupo de canais está inscrito.

3. Trabalhos Relacionados

Nesta seção, serão apresentados trabalhos similares ao que será realizado nesta monografia. Alguns deles, apenas utilizando-se de métricas diferentes para análise de comportamento em traces de mobilidade social. Cenários de mobilidade social consistem em cenários onde há a coleta de dados durante trocas de informação entre pessoas em movimento utilizando dispositivos capazes de enviar e receber dados utilizando determinada tecnologia, como, por exemplo, o *Bluetooth*.

O trabalho mostrado em [6], foi feito utilizando métricas de distância temporal para avaliar o comportamento nos traces de mobilidade. Basicamente, essa medida avaliou o nível de difusão da informação no cenário. As métricas efetivamente utilizadas nesse trabalho foram: distância temporal (menor caminho (arestas) no tempo entre o nó emissor e receptor de uma informação), métricas temporais globais (verifica a dinâmica do fluxo de informação em toda a rede) e métricas temporais locais (verifica a dinâmica do fluxo de informação em cada nó e nos seus vizinhos).

O trabalho mostrado em [7], assim como o anterior, também se utiliza de métricas temporais para avaliar e melhorar, caso possível, o comportamento de difusão de informação em cenários de mobilidade. Nesse trabalho, além de algumas métricas usadas também no trabalho anteriormente mencionado, são utilizadas outras métricas temporais, tais como: *temporal centrality*, *closeness centrality*, *temporal diameter*, entre outros. *Temporal Centrality* (Centralidade Temporal) consiste em descobrir, a cada instante temporal determinado, o nó centralizado, facilitando assim uma disseminação mais rápida da informação, aumentando a chance de a rede não “quebrar” (ausência de troca de informação). *Closeness Centrality* (Centralidade Próxima) consiste em descobrir, a cada instante temporal determinado, o número de trocas ocorridas em cada nó, descobrindo-se assim, conseqüentemente, o nó mais “popular” (o que possui um maior número de troca de informação). *Temporal diameter* (Diâmetro Temporal) consiste em descobrir, a cada instante temporal determinado, a maior distância na rede, passando por todos os nós.

O trabalho mostrado em [13], tem o enfoque em discutir sobre os diversos tipos de mobilidade humana, questionando, inclusive, sua importância. Também as classifica quanto a níveis e tipos. Dentre os tipos de níveis, existem três: estratégico, tático e operacional. O nível estratégico abrange apenas a decisão do que será feito durante o experimento (trabalhar, ir ao cinema, etc.). O nível tático abrange uma decisão estratégica sobre qual caminho irá ser traçado. O nível operacional abrange o nível da movimentação humana, como por exemplo, a velocidade dessa movimentação. Dentre os tipos de mobilidades mencionadas, estão, por exemplo: (1) a mobilidade via *Wi-Fi*, (2) via “*Inter-Vehicular Communication Protocols*” (Protocolo de Comunicação Inter-Veicular), também conhecida como VANETs e (3) via celular, utilizando uma plataforma de sinalização modularizada.

O trabalho mostrado em [14], foca a estrutura de um grafo para ilustrar os movimentos que os nós fazem durante o tempo ao trocar informação. Durante o trabalho, são mostradas fórmulas capazes de gerar algum entendimento aos estudantes, utilizando essa estrutura de um grafo, e, posteriormente, geram uma visualização gráfica dos resultados das métricas para chegarem a conclusões sobre o comportamento dos nós.

O trabalho mostrado em [19], tem como propósito o estudo das características de movimentação de dispositivos móveis e o impacto da mesma no funcionamento de uma rede móvel sem fio. Esse processo de estudo é feito via simulação. Para a realização efetiva do estudo, foram realizados cálculos de diversas medidas estatísticas. Dentre elas, está a função de distribuição cumulativa empírica (*empirical cumulative distribution function* – ECDF), que será utilizada na análise de resultados provenientes dos traces de mobilidade, nesta monografia. Em relação a alguns componentes de cada movimento capturado, como, por exemplo, velocidade dos dados reais, foi realizado um estudo comparativo entre a sua curva de distribuição cumulativa e curvas de funções de distribuição de probabilidade (normal, lognormal, exponencial, etc.). Esse tipo de estudo também será realizado nesse trabalho.

O trabalho mostrado em [20] tem o mesmo objetivo visado pelo trabalho descrito acima, porém, ao invés de simulação, o grupo de pesquisa responsável utilizou-se de oito *traces* de conectividades reais, divididos em dois grupos: *traces* mensurando contatos entre clientes e pontos de acesso, e *traces* mensurando contatos diretos entre dispositivos clientes. Dentre esses oito *traces*, três deles foram coletados em experimentos conduzidos diretamente pelo grupo de pesquisa, sendo um destes o Infocom5, utilizado mais adiante nesta monografia.

O trabalho mostrado em [21] é focado em estudar características da mobilidade e da variação do diâmetro de uma rede sem fio. A partir da manipulação de dados provenientes de três *traces* de conectividade - dois deles, Infocom5 e Infocom6, analisados também nesta monografia -, foram estudados o diâmetro das redes e a alteração do tempo de atraso, através de duas formas: removendo contatos ocorridos com certa probabilidade previamente definida e removendo contatos que tenham um tempo de duração maior ou menor que um dado valor. O objetivo era chegar a resultados que pudessem ter impacto importante na criação de algoritmos de distribuição em redes oportunísticas.

4. Investigação sobre a conectividade dos nós a partir do uso de traces de contatos sociais

As redes DTN ainda continuam longe de serem perfeitamente aplicáveis. Seu potencial é grande, mas ainda há muito estudo a ser feito para que se possa alcançar a melhor forma de implementação, se é que existirá uma única.

O primeiro passo do trabalho é a escolha dos dados de mobilidade a serem utilizados. Esses dados são provenientes da base de dados CRAWDAD (*Community Resource for Archiving Wireless Data At Dartmouth*).

Usando o *trace* dos cenários (um arquivo que mostra o instante de abertura e fechamento de todas as conexões existentes no experimento em questão), será programada uma rotina que, a partir do *trace*, fornecerá os grafos representando a conectividade da rede e medidas relevantes dessa rede. Os grafos extraídos nada mais são do que vértices representando os nós envolvidos no experimento e arestas representando os contatos estabelecidos entre esses nós, e cada um será representado, posteriormente a partir do grafo, no formato de uma matriz de adjacência. Usando as medidas calculadas, será possível tirar conclusões sobre os dados da simulação em questão.

4.1 Escolha dos dados

Como primeiro passo deste trabalho, foi escolhido o conjunto de dados de mobilidade no site da base de dados CRAWDAD. Para que o *download* dos dados fosse possível, foi preciso a criação de uma conta no site.

Em uma seção do site, há uma lista de todos os dados que podem ser baixados. Após escolher qual será utilizado, há a possibilidade da visualização de sua documentação e a devida explicação sobre como eles foram coletados, mostrando todo o contexto e facilitando a interpretação dos dados. Eles estão disponibilizados no site em diversas extensões, desde bancos de dados (.db) a simples documentos de texto (.txt).

4.2 Os Cenários Analisados

Neste trabalho, serão utilizados alguns cenários de mobilidade social (cenários onde há trocas de informação entre pessoas) para que possa ser realizada uma comparação de comportamento, dos contatos dentro dos cenários, mais eficiente, agregando assim um maior nível de conhecimento sobre esses comportamentos. A Tabela 1 mostra, resumidamente, as principais informações sobre os cenários escolhidos para esse trabalho.

| Cenários/Informações | Dispositivos | Duração |
|-----------------------------|---------------------|----------------|
| Rollernet | 62 | 3 horas |
| Infocom5 | 41 | 4 dias |
| Infocom6 | 98 | 4 dias |

Tabela 1: Dados dos cenários

Como primeiro cenário analisado, temos o *trace* “*cambridge/haggle/imote/infocom2006*”, provindo do site repositório de dados CRAWDAD. Esse *trace* consiste em um experimento realizado por aproximadamente quatro dias (do dia 24 de abril a 27 de abril de 2006), na Conferência IEEE Infocom 2006 no Princesa Sofia Gran Hotel, em Barcelona, na Espanha. Foram distribuídos os dispositivos a 70 estudantes no dia anterior ao início do experimento (dia 23 de abril) e durante esses dias os estudantes carregaram esses pequenos dispositivos *Bluetooth*,

trocando informação entre si, gerando assim os dados contidos nesse *trace*. Em [21], os pesquisadores destacam que os dados relacionados à duração dos contatos podem não ser muito precisos, devido a problemas como intervalo entre buscas *bluetooth*, limitações de hardware e interferência externa, podendo refletir em contatos sendo reportados como mais curtos do que realmente ocorreram e até mesmo contatos ignorados.

Como segundo cenário analisado, temos o *trace* “*upmc/rollernet (v. 2009-02-02)*”, também provindo do repositório CRAWDDAD. Esse *trace* consiste de um experimento realizado, em 20 de agosto de 2006, por patinadores em um *tour* em Paris, na França. Foram distribuídos 62 aparelhos *iMote*, e por aproximadamente três horas, dados trafegados com a tecnologia *Bluetooth* foram coletados. O trabalho que originou esse *trace* ressalta que, por usar tecnologia *Bluetooth*, era possível que dois aparelhos que estivessem fazendo uma busca ao mesmo tempo não se enxergassem, mesmo estando dentro do alcance um do outro. Para contornar esse problema, os tempos de contatos entre os *iMotes* foram “completados” da seguinte forma: caso um contato entre dois *iMotes* seja observado num instante t , mas não em $t+1$ e então novamente em $t+2$, assume-se que o contato nunca deixou de existir, e com isso teríamos uma duração de contato de três segundos.

Como terceiro cenário analisado, temos outro *trace* provindo do CRAWDDAD, consistindo do já citado experimento realizado na Conferência IEEE Infocom 2005, evento que ocorreu em Miami, nos EUA [20]. Foram distribuídos 41 dispositivos aos participantes para a coleta dos dados e formação do cenário a ser analisado, ao longo de quatro dias. No trabalho original, foram considerados para a coleta de dados dois tipos de contato: contatos “internos”, aqueles ocorridos entre os 41 dispositivos distribuídos, e contatos “externos”, quando algum dos dispositivos detectava um dispositivo *bluetooth* que não pertencia a esse grupo. Nesta monografia, trataremos apenas dos contatos internos.

5. Metodologia utilizada para análise da conectividade

5.1 Rotina de transformação

Com os dados escolhidos, é necessária agora uma rotina que gere, a partir do *trace* dos dados de mobilidade, grafos que ilustram o comportamento dos nós a cada instante. A partir desses grafos, serão geradas matrizes de adjacências, a fim de facilitar a obtenção de medidas necessárias para um uso futuro dentro deste trabalho.

A linguagem de programação utilizada para a criação desta rotina foi Java, e os ambientes utilizados para a programação foram o *Netbeans* e, posteriormente, o *Eclipse*.

Os *traces* utilizados contêm informações apenas acerca das conexões estabelecidas durante a simulação original, sendo formatado com cinco colunas: a primeira contém uma contagem de tempo em segundos; a segunda contém o termo '*CONN*', indicando que a informação é sobre uma conexão; a terceira e a quarta indicam os nós envolvidos; e a quinta coluna contém os termos '*up*' ou '*down*', dependendo se naquele instante a conexão está sendo estabelecida ou exterminada.

Dessa forma, a rotina desenvolvida faz uma leitura de cada uma das linhas do arquivo do *trace*, identificando os instantes em que cada contato começa e termina, primeiramente armazenando em uma lista o nome de cada nó participante da simulação. Posteriormente, cria um novo arquivo armazenando essa informação, que será utilizado para a criação das matrizes de adjacência. Também é criado um terceiro arquivo, que, ao invés de guardar os instantes de início e fim do contato, armazena apenas o tempo de duração do mesmo. Este terceiro arquivo será utilizado para a métrica de distribuição das durações dos contatos.

Com as informações contidas no primeiro arquivo, será criado, para cada instante, a estrutura de grafo correspondente aos nós e suas conexões no determinado instante. Após isso, será gerada uma matriz de adjacência a partir de cada grafo gerado. Cada dispositivo distribuído às pessoas em cada um dos cenários representa um nó. As matrizes serão quadradas, contendo como linhas e colunas o número correspondente a todos os nós (dispositivos) da rede simulada – não apenas aqueles que tiverem contato estabelecido naquele segundo específico -, e serão criadas completamente zeradas para apenas em seguida serem devidamente preenchidas.

Um *loop* com número de iterações igual ao número de segundos de duração do *trace* será responsável por preencher devidamente os contatos entre os nós em cada uma das matrizes, identificando os dois nós e preenchendo na matriz, em duas posições (onde o primeiro nó é linha e o segundo é coluna, e vice-versa), com o valor “1”, identificando a existência de uma conexão entre os valores de linha e coluna da matriz. O preenchimento igual nas posições simétricas (ex: [0] [1] e [1] [0]) se deve ao fato de esse grafo não ser direcional (dígrafo); caso fosse um dígrafo, a posição simétrica não necessariamente possuiria o mesmo valor. A Figura 3 mostra um exemplo de uma dessas matrizes geradas.

| | p0 | p1 | p2 | |
|----|----|----|----|---|
| p0 | | 0 | 1 | 1 |
| p1 | | 1 | 0 | 1 |
| p2 | | 1 | 1 | 0 |

Figura 3: Exemplo de matriz de adjacência

5.2 Geração das medidas

Ao final da rotina descrita, teremos uma matriz de adjacência ilustrando cada um dos instantes (em segundos) observados no trace. Cada uma delas representa um grafo G_t , onde t indica o instante da simulação representado pelo grafo, montado utilizando os dispositivos sem fio distribuídos durante o experimento como seus vértices e os contatos estabelecidos entre eles como arestas.

Para iniciar o processo de escolha de quais medidas, baseadas na estrutura de um grafo, seriam implementadas, foi pensado em que métricas seriam realmente fundamentais no contexto e que poderiam agregar um bom conhecimento e, conseqüentemente, uma boa conclusão.

Após esse processo de análise, foram escolhidas as seguintes medidas: grau médio, densidade, grau no tempo, coeficiente de agrupamento, tempo de conexão e duração dos contatos. A seguir, será comentada a importância dessas métricas no contexto das redes DTN no âmbito social.

A densidade [4] é matematicamente definida pela equação 5.1 a seguir:

$$D_{G(A,N)_t} = \frac{2C}{X(X-1)}, \quad C > 0, \quad X \geq 2, \quad t > 0; \quad (5.1)$$

onde D é a densidade do grafo G , C é o número total de arestas contidas no conjunto A e pertencentes ao grafo e X é o número total de nós contidos no conjunto N e pertencentes ao grafo. A métrica da densidade verifica o nível da interação mediante todas as possibilidades encontradas na rede, relacionando o número de arestas com o número possível de arestas, observado através da fórmula $X(X-1)$. Quanto maior for o valor da densidade, mais densa será essa rede, havendo maior índice de troca de informação entre os nós. Quanto menor, mais esparsa será a rede, havendo um menor índice de interação entre os nós.

O grau g de um nó é numericamente representado pelo número de arestas conectadas a ele. A métrica do grau médio, intimamente relacionada à da densidade, verifica, primeiramente, o nível de interação de cada nó presente na rede e, posteriormente, calcula o nível de interação geral da rede, através de uma média, sendo obtida através da equação 5.2:

$$gm_t = \frac{\sum_{i=0}^N g_i}{X}, \quad t > 0, \quad i \in N \quad (5.2)$$

Calcula-se o grau médio de um dado instante t da simulação primeiramente somando o grau de cada nó i , para em seguida dividir o resultado dessa soma pelo número total de nós X .

Usando o conceito de matriz de adjacência, podemos chegar a esse resultado da seguinte forma: percorre-se as linhas e as colunas, incrementando um contador a cada vez que uma interseção linha/coluna contenha um valor diferente de zero – representando um contato -, e quando termina-se de percorrer a matriz, divide-se o valor obtido pelo número de linhas, obtendo assim um valor médio de arestas que saem dos nós da rede, o grau médio. Esse valor é importante para se ter uma noção do número de trocas de informação que se pode haver durante a simulação. Quanto maior esse valor, maior a tendência de haver mais conexões entre os nós.

Vale ressaltar que o grau médio nem sempre é uma métrica confiável. O nível de confiabilidade dessa métrica depende muito das características do próprio cenário. Por exemplo, usando os cenários Infocom utilizados na monografia, caso haja a necessidade de calcular o grau médio em um intervalo de tempo específico, e nesse intervalo não houver a realização da conferência, o resultado da métrica não irá ilustrar de maneira correta o comportamento dessa métrica no cenário como um todo.

O grau de um nó i no tempo é responsável por verificar o comportamento de um determinado nó durante o tempo da simulação, mostrando suas mudanças de conexão. Essa medida tem o intuito de verificar a intensidade com que o nó entra em contato com outros nós, sendo no âmbito temporal (longos contatos) ou no âmbito quantitativo (número total de contatos). Ao final, fazendo essa análise em todos os nós, será conhecido o mais suscetível ao contato.

O coeficiente de agrupamento (*clustering coefficient*) [3] de um nó i , é a razão entre o número de arestas existentes entre os vizinhos de i e o número máximo de arestas possíveis entre estes vizinhos. Um nó é vizinho do nó i quando este possui um contato direto (aresta) com o nó i . Tal razão representa uma medida da densidade de arestas estabelecidas entre os vizinhos de um nó. O coeficiente de agrupamento de uma rede é calculado como a média dos coeficientes de agrupamento de todos os seus nós. O coeficiente de agrupamento de um nó é obtido através da equação 5.3:

$$C_i = \frac{2 a_i}{k_i (k_i - 1)}, \quad a \in A, \quad i \in N \quad (5.3)$$

onde k_i é o número de nós vizinhos a i e a_i é o número de conexões entre esses vizinhos. Esse conceito de coeficiente de agrupamento é mais facilmente verificado através da noção de lista de adjacência: toma-se um nó específico, e para cada nó vizinho, verifica-se a lista de adjacências referentes ao mesmo, comparando com a lista do primeiro nó. A cada vez que um mesmo nó aparecer como vizinho em ambas as listas em comparação, será incrementada a contagem A de arestas; ao fim dessa contagem, esse valor será dividido pelo número máximo de ligações possíveis entre esses nós, o que é obtido pela fórmula $\frac{k_i(k_i - 1)}{2}$.

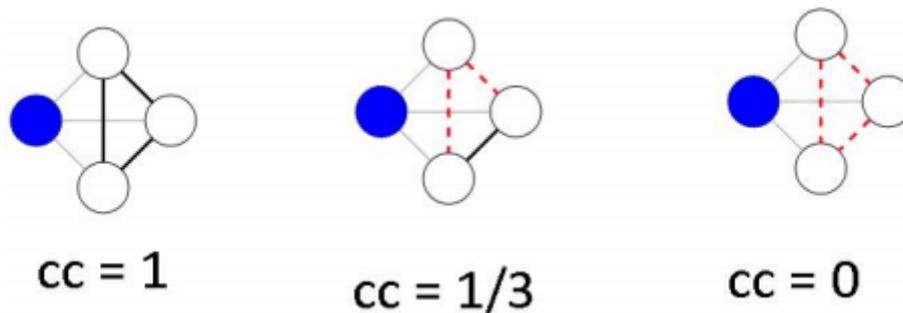


Figura 4: Exemplo do cálculo do coeficiente de clusterização (fonte: [3])

A ilustração fornecida pela Figura 4 ajuda no entendimento teórico do coeficiente de agrupamento: tomamos o nó destacado em azul como o nó i , sendo assim os nós brancos seus vizinhos. A primeira imagem mostra uma situação onde os 3 nós vizinhos de i se conectam entre si, fazendo com que sejam estabelecidos 3 contatos dentre os 3 possíveis, fazendo com que o coeficiente de agrupamento de i seja 1; analogamente, a segunda imagem mostra apenas um contato sendo estabelecido dentre os 3 possíveis, e com isso o coeficiente de agrupamento é de 1 em 3; já a terceira imagem mostra uma situação em que não há contatos entre os nós vizinhos de i , e portanto o valor do coeficiente de agrupamento é 0.

A duração de conexão verifica o comportamento de cada conexão durante o tempo de simulação, podendo proporcionar um conhecimento sobre quais nós se conectaram mais vezes no tempo, bem como quais contatos duraram mais tempo. A forma como essa métrica é adquirida já foi explicada na subseção anterior: o arquivo do *trace* é lido e é criada uma lista contendo em cada linha três colunas: as primeiras duas contendo os nós envolvidos no contato e a terceira contendo o tempo de duração do contato.

A distribuição da duração dos contatos nos mostra o número de ocorrências que cada tempo de duração teve, em contatos. A partir da lista gerada na métrica da duração de conexão, é gerado um mapa contendo uma única vez cada um dos diferentes tempos de duração encontrados na segunda coluna da lista de durações e o número de ocorrências de cada um desses tempos, começando com zero. Então, um *loop* percorre novamente a primeira lista e incrementa no mapa o número de ocorrências de cada tempo, conforme for aparecendo. Ao final do *loop*, teremos os tempos de duração armazenados de forma ordenada, associados ao número de contatos que duraram tal tempo; a relação 1-200, por exemplo, mostrará que 200 contatos tiveram duração de 1 segundo.

O conhecimento sobre a distribuição da duração dos contatos é importante, por exemplo, pra quem quiser desenvolver um novo protocolo de roteamento para esse tipo de rede. Se, em geral, os contatos forem de curta duração, o protocolo terá que atuar de maneira rápida e eficaz, caso contrário, não haverá a necessidade de ser rápido.

Ao final da execução da rotina de transformação dos dados em matrizes de adjacência e do cálculo das métricas, serão gerados arquivos com o resultado dessas métricas. Dependendo da métrica, ela pode gerar um ou mais arquivos.

6. Resultados obtidos

Nesta seção, as métricas, calculadas a partir dos *traces* e execução de suas rotinas geradoras, serão analisadas e, posteriormente, a partir desta análise, serão apresentadas as conclusões tiradas sobre o comportamento de cada métrica de cada *trace* estudado.

Para facilitar a análise dos dados, será utilizado o *software* livre R, o qual nos possibilita gerar gráficos a partir desses arquivos e analisar o comportamento das métricas nesses gráficos.

Uma das análises consiste em, para cada métrica implementada, gerar seu gráfico ECDF (*Empirical Cumulative Distribution Function*) [19], que consiste em mostrar o comportamento da curva de sua função de distribuição acumulativa. A partir desse gráfico, o próximo passo será gerar diversas distribuições de probabilidade (normal, lognormal, exponencial e Weibull) e, através de um teste de aderência, descobrir a qual dessas distribuições estatísticas a curva de distribuição acumulativa da métrica mais se assemelha. Para cada métrica, também será gerado um histograma, sendo uma outra forma de analisar os dados.

Para a geração das curvas, cada distribuição possui seus próprios parâmetros. A distribuição Normal possui os seguintes parâmetros: *mean* (média dos valores) e *sd* (desvio padrão dos valores). A distribuição Lognormal possui os seguintes parâmetros: *meanlog* (média na escala logarítmica dos valores) e *sdlog* (desvio padrão na escala logarítmica dos valores). A distribuição Exponencial tem como único parâmetro o *rate* (taxa). A distribuição Weibull possui os seguintes parâmetros: *shape* (forma) e *scale* (escala).

O teste de aderência a ser utilizado neste trabalho será o *Mean Squared Error* (MSE), ou Erro Quadrático Médio, que calcula a diferença entre cada ponto entre dois parâmetros (no caso desse trabalho, entre duas curvas), até chegar a um valor médio de

erro. Esse valor médio é o que será utilizado para determinar que distribuições são mais aderentes a que métricas.

Existem outros testes de aderência, como, por exemplo, o teste *Kolmogorov-Smirnov*. A justificativa da utilização do MSE se deve à simplicidade de seu uso e o entendimento de seu procedimento de cálculo e resultado gerado.

Os resultados das análises são mostrados em subseções, dividindo-os de acordo com os cenários escolhidos.

6.1 Rollernet

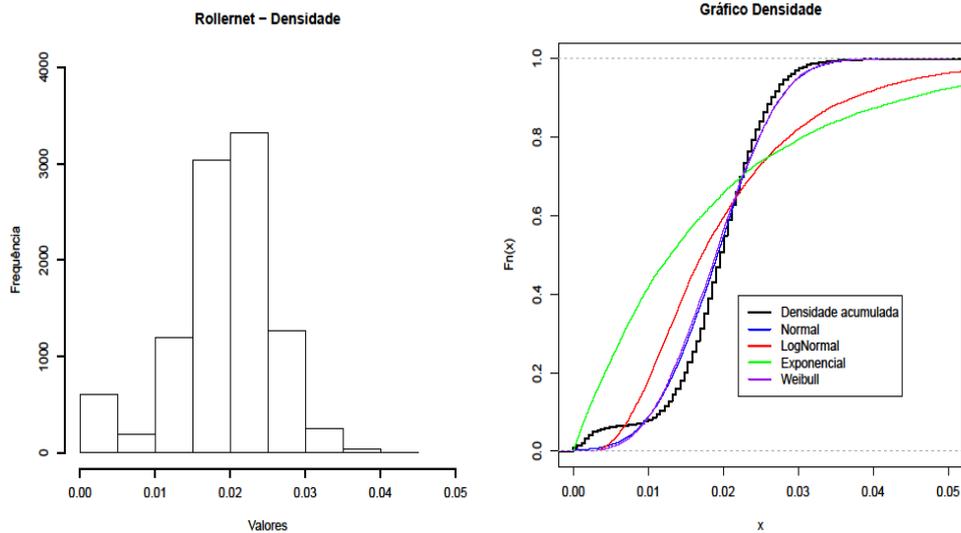
Como mencionado anteriormente, o Rollernet consistiu em um evento de patinadores em Paris, onde cada um carregava um aparelho *IMote* para a realização do tráfego de dados. Esse experimento durou, aproximadamente, 3 horas.

A Tabela 2 mostra os dados estatísticos das métricas no cenário Rollernet. Esses dados foram retirados baseados nas curvas ECDF de cada métrica. Pode-se notar que existem alguns valores em verde. Esses valores são os menores de cada métrica, o que significa que a distribuição estatística correspondente à coluna é a mais aderente à curva ECDF da métrica.

| Métricas/Distribuições | Normal | Lognormal | Exponencial | Weibull |
|--|---------------|--------------|--------------|---------------|
| Densidade | 0.00008814746 | 0.0002267156 | 0.0004202215 | 0.00008743567 |
| Grau Médio | 0.0819699 | 0.21023 | 0.3908432 | 0.08129656 |
| Coefficiente de Clusterização Médio | 0.00318931 | 0.004145588 | 0.004484768 | 0.002965294 |
| Frequência da duração de contatos no tempo | 0.0002573005 | 0.0002307354 | 0.0001638353 | 0.0002456629 |

Tabela 2: Resultados do MSE das métricas do cenário Rollernet

Como se pode ver, a densidade nesse cenário possui uma curva cumulada com um comportamento mais aderente à distribuição Weibull, sendo que a distribuição Normal também apresenta uma curva de comportamento bem semelhante, com apenas uma pequena diferença para mais.



(a) Histograma

(b) ECDF

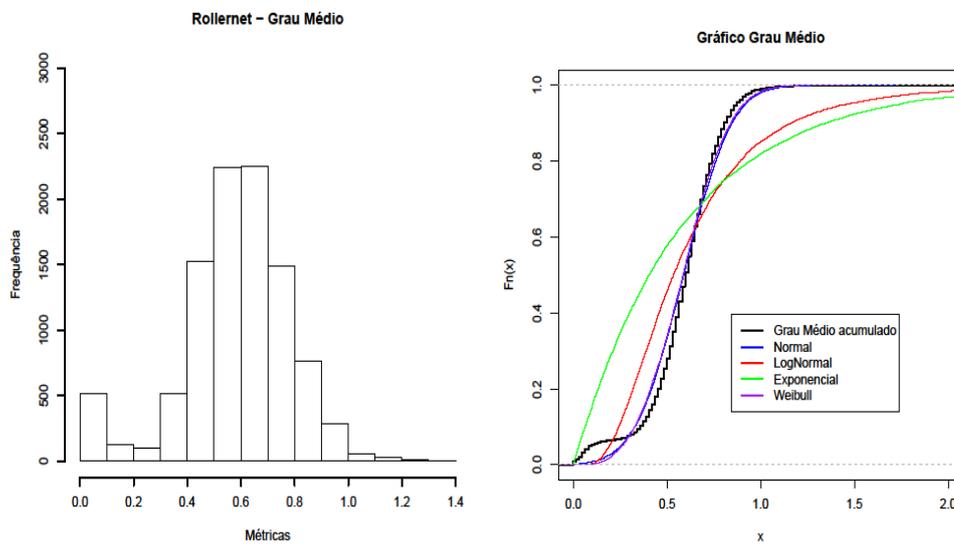
Figura 5: Análise da Densidade no cenário Rollernet

O histograma da Figura 5(a) nos mostra que o comportamento da variação dos resultados calculados da densidade durante o tempo tem uma forma muito parecida com as distribuições Normal e Weibull, tendo seu ápice de quantidade de valores na parte central do gráfico e valores menos recorrentes nas extremidades. A média dos valores, por consequência, também se encontra na parte central do gráfico, sendo igual a, aproximadamente, 0.019 e possuindo um desvio padrão igual a, aproximadamente, 0.007.

Dado que a coleta de dados do cenário Rollernet possui uma duração de 9.939 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente 3.000 vezes, ou seja, em 3.000 instantes (segundos), conclui-se que o valor médio da métrica da densidade nesse cenário ocorreu em, aproximadamente, 33% do tempo total dele.

Observando a Figura 5 (b), tem-se a análise gráfica comprovando a análise numérica. Percebe-se que as duas curvas (distribuições Normal e Weibull) estão, quase em totalidade de tempo, sobrepostas, o que justifica a pequena diferença numérica entre os valores.

Quanto ao grau médio, a Tabela 2 mostra uma situação numérica equivalente à observada na densidade. Os valores relacionados à distribuição Normal e Weibull possuindo uma diferença numérica muito pequena, tendo novamente a distribuição Weibull como a mais aderente à curva ECDF da medida.



(a) Histograma

(b) ECDF

Figura 6: Análise do Grau Médio no cenário Rollernet

A Figura 6(a) mostra um comportamento semelhante ao ocorrido na densidade. Novamente o ápice de frequência se encontra na parte central do gráfico e as menores ocorrências nas extremidades, sendo desta vez a média das frequências sendo igual a, aproximadamente, 0.58 e desvio padrão igual a, aproximadamente, 0.2.

Dado que a coleta de dados do cenário Rollernet possui uma duração de 9.939 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 2.250 instantes, conclui-se que o valor médio da métrica do grau médio nesse cenário ocorreu em, aproximadamente, 23% do tempo total dele.

A Figura 6(b) ilustra realmente o comportamento numérico mostrado na Tabela 2, tendo as curvas das distribuições Weibull e Normal sobrepostas praticamente em todo momento do gráfico, justificando a pequena diferença numérica nos valores do resultado do teste MSE.

Ainda no conceito de grau médio, foi realizada mais uma análise. Utilizando o *trace* do cenário Rollernet, foram coletados três nós mediante suas frequências dentro de todas as conexões do cenário. Os três nós foram: “p23”, por ser o nó com o menor número de contatos; “p49” por ser o nó de maior número de contatos e “p55” por ser o nó que se conectou um número de vezes mais próximo à quantidade média de conexões. Foram geradas as curvas ECDF relacionadas à variação do grau de cada nó em cada instante da coleta de dados e utilizou-se a curva ECDF do grau médio usada anteriormente como parâmetro, para averiguar qual nó possui uma variação de grau mais aderente à do grau médio. A Tabela 3 mostra os resultados estatísticos dessa análise.

| Nós/ Medida | Grau Médio |
|-------------|------------|
| p23 | 0.6536343 |
| p49 | 5.503577 |
| p55 | 1.591812 |

Tabela 3: Resultados do MSE da variação do grau dos nós no cenário Rollernet

Como se pode ver na Tabela 3, o grau “p23”, o de menor número de contatos no *trace*, possuiu um comportamento de variação mais aderente ao da variação do grau médio. Nota-se também que a diferença em relação aos outros nós foi grande, principalmente ao nó de maior número de contatos, o “p49”.

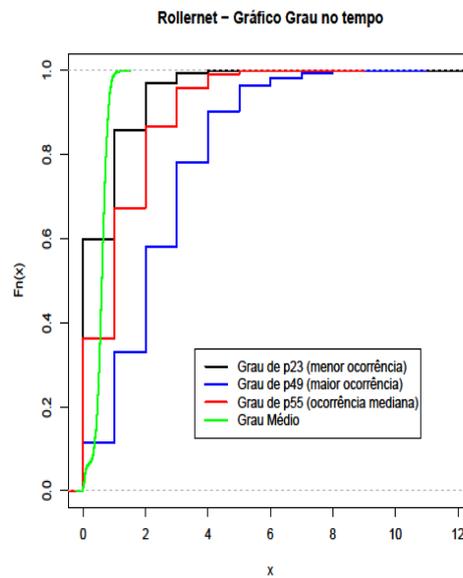


Figura 7: Análise de Variação do Grau dos nós no cenário Rollernet

A Figura 7 ilustra o comportamento mostrado pela Tabela 3, onde vemos a curva preta (a do nó “p23”) mais próxima da curva em verde, que é a curva referente à ECDF do grau médio. Isso reafirma que, no geral, o grau médio de cada instante foi um valor muito baixo, mais próximo do valor do grau do nó menos conectado, o que foi mostrado na análise anterior, onde chegou-se à conclusão de que foi em torno de 0,6, ou seja, em cada instante, muitos nós não realizavam contatos.

Quanto ao coeficiente de agrupamento (ou coeficiente de clusterização), a Tabela 2 mostra que, assim como nas outras métricas, as distribuições Normal e Weibull são as duas mais aderentes, sendo, novamente, a distribuição Weibull a mais aderente à curva ECDF do coeficiente de agrupamento.

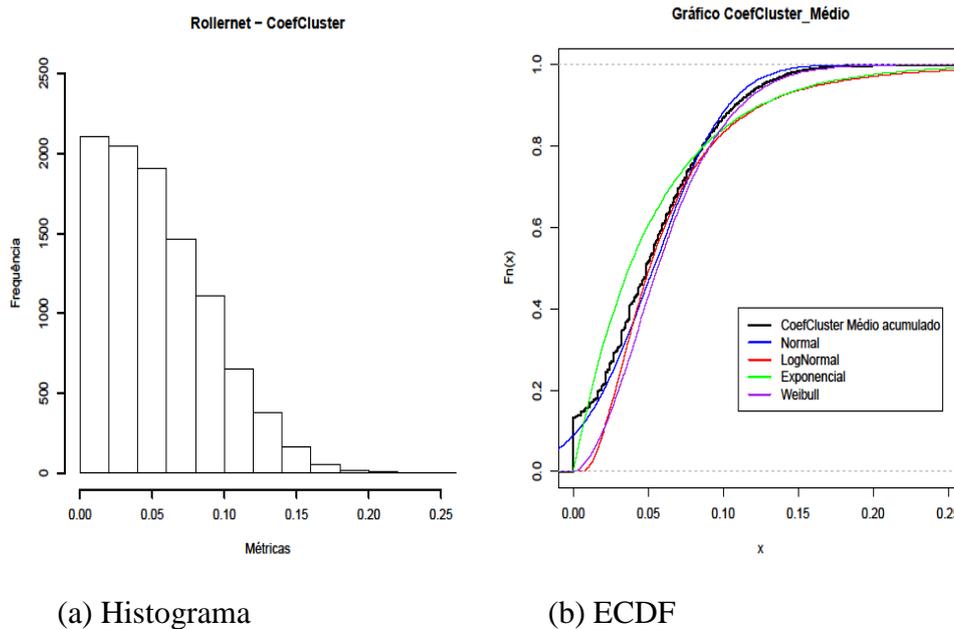


Figura 8: Análise do Coeficiente de Agrupamento no cenário Rollernet

O histograma acima, mostrado na Figura 8(a), ilustra a variação dos valores do coeficiente de agrupamento calculados a cada instante da coleta de dados. Ela mostra um comportamento diferente dos vistos anteriormente. Nesse caso, os ápices de frequência de valores se encontram na parte esquerda do gráfico, simulando uma “metade final” do desenho das distribuições Normal e Weibull. Esse fato mostra que a média de valores de coeficiente de agrupamento médio por instante é baixo, ou seja, em geral, as conexões são espaçadas, não concentradas em um mesmo grupo de nós. No caso, a média do coeficiente de agrupamento médio em cada instante é igual, aproximadamente, a 0.05 e seu desvio padrão é igual, aproximadamente, a 0.04.

Dado que a coleta de dados do cenário Rollernet possui uma duração de 9.939 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 2.000 instantes, conclui-se que o valor médio da métrica do coeficiente de agrupamento nesse cenário ocorreu em, aproximadamente, 20% do tempo total dele.

A Figura 8(b) mostra, graficamente, o que a Tabela 2 mostra estatisticamente. Note que, nesse caso, as curvas das distribuições estão todas mais próximas, o que se percebe pela diferença entre os valores mostrados na tabela, menores que nas duas métricas anteriores. Mas percebe-se também que, na parte entre os valores 0,6 e 1,0 do eixo $F_n(x)$, a curva da distribuição Weibull é a que mais se aproxima da curva ECDF da métrica.

Quanto à frequência de duração dos contatos, os dados calculados foram normalizados antes da análise, ou seja, foram colocados no intervalo entre zero e 1. A Tabela 2 nos mostra que a curva ECDF dessa métrica é mais aderente à distribuição Exponencial, diferente das outras métricas. Mas, nesse caso, as quatro distribuições também possuem uma diferença numérica pequena entre si.

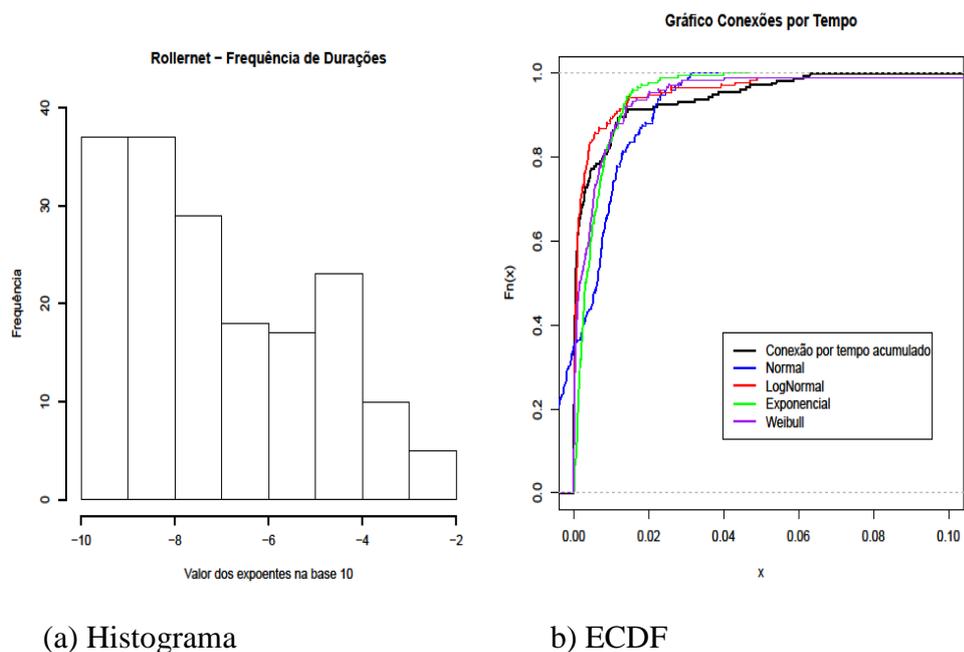


Figura 9: Análise da Frequência de Duração dos Contatos no cenário Rollernet

O histograma acima, mostrado na Figura 9(a), ilustra o comportamento da variação dos resultados calculados e normalizados durante o tempo de coleta de dados. Este gráfico possui os valores de seu eixo 'x' em escala logarítmica. Cada valor representa um expoente, e todos se encontram na base 10. A Figura 9(a) mostra que o ápice da frequência dos resultados encontra-se na parte esquerda do gráfico, mostrando que a média deles é baixa, o que permite concluir que, em sua grande maioria, os contatos estabelecidos entre os nós são de curta duração. A média dessa métrica nesse cenário é igual a, aproximadamente, 0.006 e o desvio padrão é igual a, aproximadamente, 0.01.

Dado que a coleta de dados do cenário Rollernet possui uma quantidade total de 15.803 contatos, e a média dos valores ocorreu, segundo o histograma, aproximadamente 150 vezes, conclui-se que o valor médio da métrica da frequência da duração dos contatos nesse cenário ocorreu em, aproximadamente, 1% da quantidade total de nós dele.

A Figura 9(b) mostra exatamente o que foi concluído observando os valores da Tabela 2: as quatro distribuições juntas na maior parte do gráfico. Em casos como este,

somente os testes de aderência podem mostrar estatisticamente qual distribuição é a mais aderente.

6.2 Infocom5

Como mencionado anteriormente, o Infocom5 foi um evento ocorrido em Miami que durou quatro dias, onde foram distribuídos 41 dispositivos a participantes para a realização do tráfego de dados.

A Tabela 4, a seguir, mostra os resultados estatísticos da análise das métricas dentro desse cenário. Novamente, os valores em verde mostram que a distribuição referenciada na coluna é a mais aderente à curva ECDF da métrica relacionada na linha. Note que o teste de aderência entre a curva da distribuição Weibull e a curva ECDF da métrica de frequência da duração de contatos no tempo não foi realizado. Isso se deve a problemas de propriedade entre a distribuição e a métrica calculada.

| Métricas/Distribuições | Normal | Lognormal | Exponencial | Weibull |
|--|--------------|--------------|--------------|---------------|
| Densidade | 0.0002894823 | 0.0002667159 | 0.0003102623 | 0.0002524406 |
| Grau Médio | 0.1158139 | 0.1065693 | 0.1240965 | 0.1008075 |
| Coefficiente de Clusterização Médio | 0.004298606 | 0.00488861 | 0.003992625 | 0.004566544 |
| Frequência da duração de contatos no tempo | 0.0003943468 | 0.0002055635 | 0.0002057899 | Não calculado |

Tabela 4: Resultados do MSE das métricas do cenário Infocom5

Como se pode ver, com relação à densidade, sua curva ECDF possui a da distribuição Weibull como a curva mais aderente, mas os valores em geral possuem uma diferença numérica bem pequena entre si.

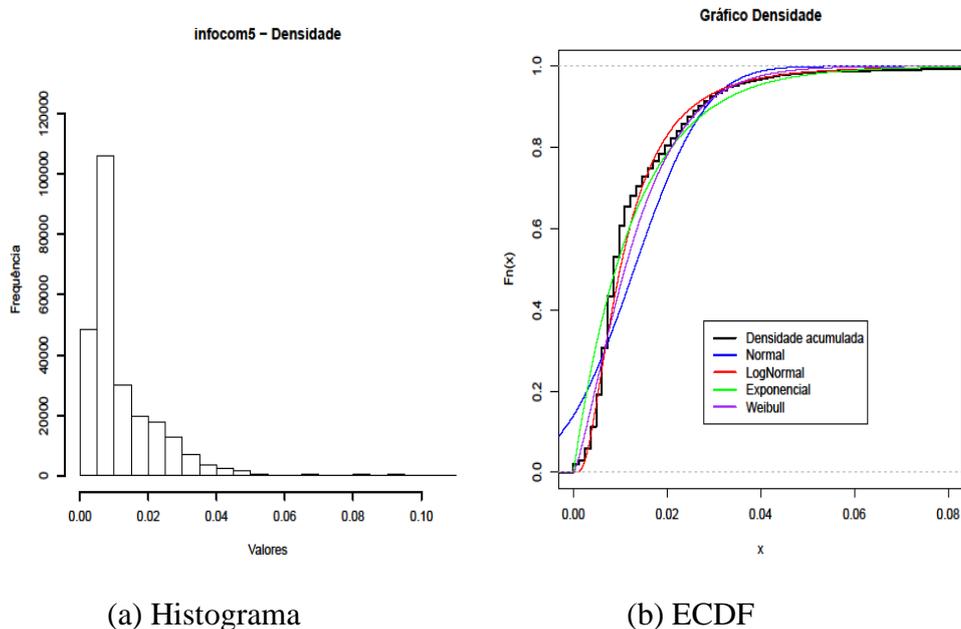


Figura 10: Análise da Densidade no cenário Infocom5

O histograma, mostrado acima pela Figura 10(a), ilustra o comportamento da variação dos resultados calculados da métrica em cada instante do cenário. Ela mostra que o ápice da frequência dos dados se encontra na parte esquerda do gráfico, o que permite concluir que os valores da densidade a cada instante no cenário, em geral, foram pequenos, ou seja, o índice de conexões simultâneas em um mesmo instante foi baixo. A média dos valores, nesse caso, é de, aproximadamente, 0,01 e o desvio padrão igual a, aproximadamente, 0,01.

Dado que a coleta de dados do cenário Infocom5 possui uma duração de 254.150 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 100.000 instantes, conclui-se que o valor médio da métrica da densidade nesse cenário ocorreu em, aproximadamente, 39% do tempo total dele.

Como se pode ver, a Figura 10(b) ilustra claramente o que a Tabela 4 mostra estatisticamente. O fato de as curvas das quatro distribuições estarem próximas justifica a pequena diferença entre seus resultados no teste de aderência. Também é perceptível

que a curva da distribuição Weibull é a que mais se aproxima da curva da métrica, principalmente no intervalo entre os valores 0,8 e 1,0 do eixo $F_n(x)$.

Quanto ao grau médio, a Tabela 4 nos mostra que a curva da distribuição Weibull é a mais aderente à curva ECDF do grau médio. Novamente, a diferença entre os valores foi pequena, sendo a distribuição Lognormal a que mais se aproximou da distribuição Weibull.

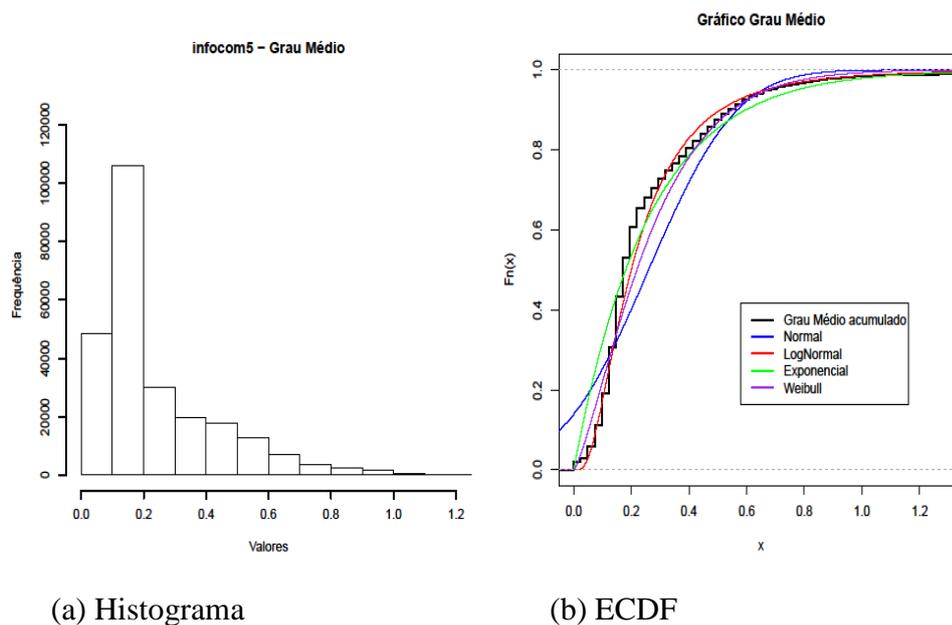


Figura 11: Análise do Grau Médio no cenário Infocom5

O histograma, mostrado acima na Figura 11(a), ilustra o comportamento da variação dos resultados calculados ao longo da coleta dos dados. Ela mostra que o ápice da frequência dos dados se encontra na parte esquerda do gráfico, o que diz que, em geral, os graus médios dos nós a cada instante foram baixos, ou seja, poucos nós participaram de múltiplas conexões simultâneas. A média de ocorrência do grau médio nesse cenário foi de, aproximadamente, 0,26 e o desvio padrão igual a, aproximadamente, 0,24.

Dado que a coleta de dados do cenário Infocom5 possui uma duração de 254.150 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em

110.000 instantes, conclui-se que o valor médio da métrica do grau médio nesse cenário ocorreu em, aproximadamente, 43% do tempo total dele.

A Figura 11(b) mostra as quatro curvas das distribuições bem juntas, novamente justificando a pequena diferença entre os resultados estatísticos dos testes de aderência. Mas percebe-se ao longo do gráfico que a curva da distribuição Weibull é a que mais se aproxima da curva da métrica, principalmente entre os valores 0,8 e 1,0 do eixo $F_n(x)$, onde quase se sobrepõem.

Neste cenário também foram coletados três nós para a análise comparativa com a curva ECDF do grau médio. Os três graus foram: “p30”, por ser o nó com o menor número de contatos; “p32”, por ser o nó que se conectou um número de vezes mais próximo à quantidade média de conexões; e “p39”, por ser o nó de maior número de contatos. A Tabela 5, a seguir, mostra os resultados estatísticos dos testes de aderência entre as curvas ECDF da variação dos graus desses nós no tempo de coleta e a curva ECDF do grau médio deste cenário.

| Nós/ Medida | Grau Médio |
|--------------------|-------------------|
| p30 | 0.2847142 |
| p32 | 1.927733 |
| p39 | 1.955852 |

Tabela 5: Resultados do MSE da variação do grau dos nós no cenário Infocom5

A Tabela 5 mostra que o nó de menor ocorrência é o que possui a curva de variação mais aderente à curva ECDF do grau médio. Possui também uma grande diferença numérica em relação aos outros nós. A Figura 18 dará uma ideia melhor sobre o que esses números mostram.

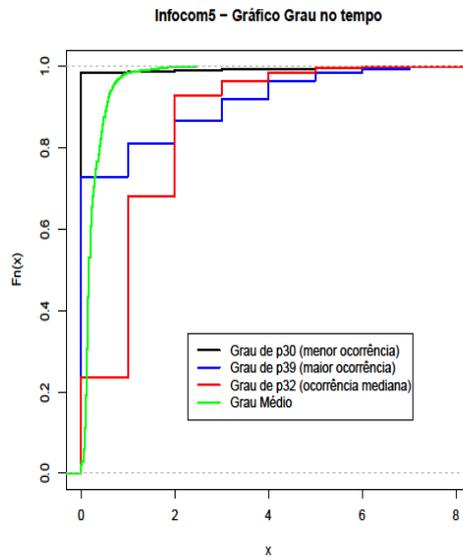


Figura 12: Análise de Variação do Grau dos nós no cenário Infocom5

A Figura 12 mostra a curva referente ao nó de menor ocorrência sendo a mais próxima da curva ECDF referente ao grau médio. No ponto 1.0 do eixo $F_n(x)$ elas se sobrepõem, enquanto as outras se mantêm longe, justificando a diferença alta entre seus valores numéricos. Este fato é mais uma coisa que ressalta os valores baixos do grau médio nesse cenário, cuja média foi concluída em 0,26 na análise anterior.

Quanto ao coeficiente de agrupamento, a Tabela 4 mostra que a curva da distribuição Exponencial é a mais aderente à curva ECDF do coeficiente, sendo a distribuição Normal a segunda mais aderente. Este fato mostra que o coeficiente de agrupamento médio, nesse cenário, tem um nível de crescimento, de um instante a outro, maior que o das outras métricas.

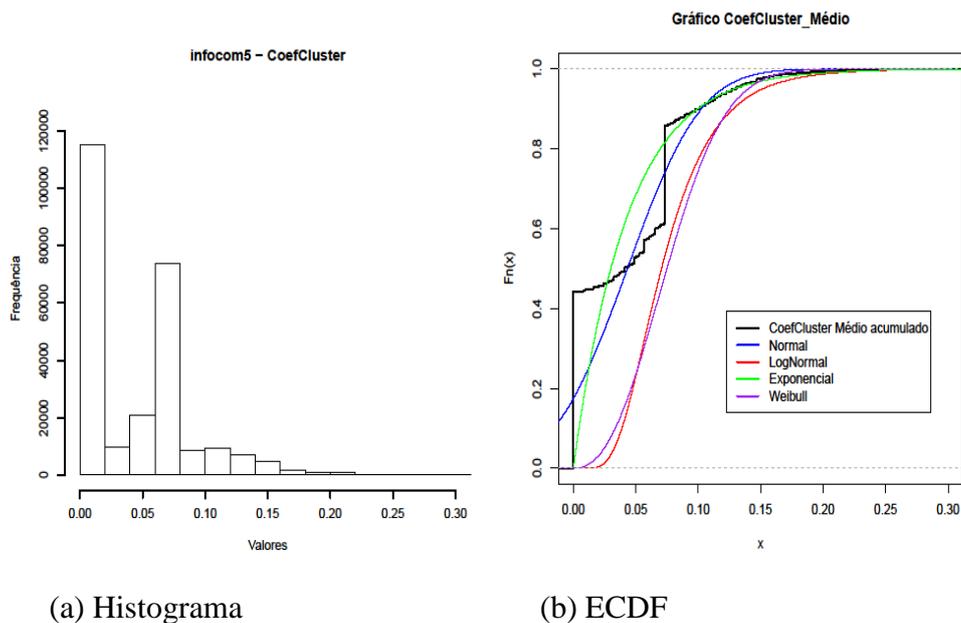


Figura 13: Análise do Coeficiente de Agrupamento no cenário Infocom5

O histograma, mostrado na Figura 13(a), ilustra o comportamento da variação dos valores calculados do coeficiente de agrupamento médio a cada instante da coleta dos dados. Ela mostra o ápice de frequência sendo, novamente, na parte esquerda do gráfico, ilustrando valores baixos de coeficiente na maioria dos instantes. Diferente nesse gráfico é um segundo pico, menor que o outro, também mais à esquerda do gráfico, aproximadamente no valor 0,07. Nesse cenário, a média da métrica está entre esses dois picos, em aproximadamente, 0,04 e o desvio padrão igual a, aproximadamente, 0,05.

Dado que a coleta de dados do cenário Infocom5 possui uma duração de 254.150 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 120.000 instantes, conclui-se que o valor médio da métrica do coeficiente de agrupamento nesse cenário ocorreu em, aproximadamente, 47% do tempo total dele.

A Figura 13(b) mostra a curva da distribuição Exponencial, na maior parte do gráfico, como a mais próxima da curva ECDF do coeficiente. Na parte baixa do eixo $F_n(x)$, até o valor 0,2, a distribuição Normal estava também próxima, mas na parte de cima, depois do valor 0,8, a curva Exponencial se sobrepõe à curva da métrica.

Quanto a frequência de duração dos contatos, a Tabela 4 mostra que a distribuição Lognormal foi a curva mais aderente à curva ECDF da frequência. Percebe-se também que a distribuição Exponencial também se aproximou, tendo uma diferença bem pequena para mais em relação à distribuição Lognormal.

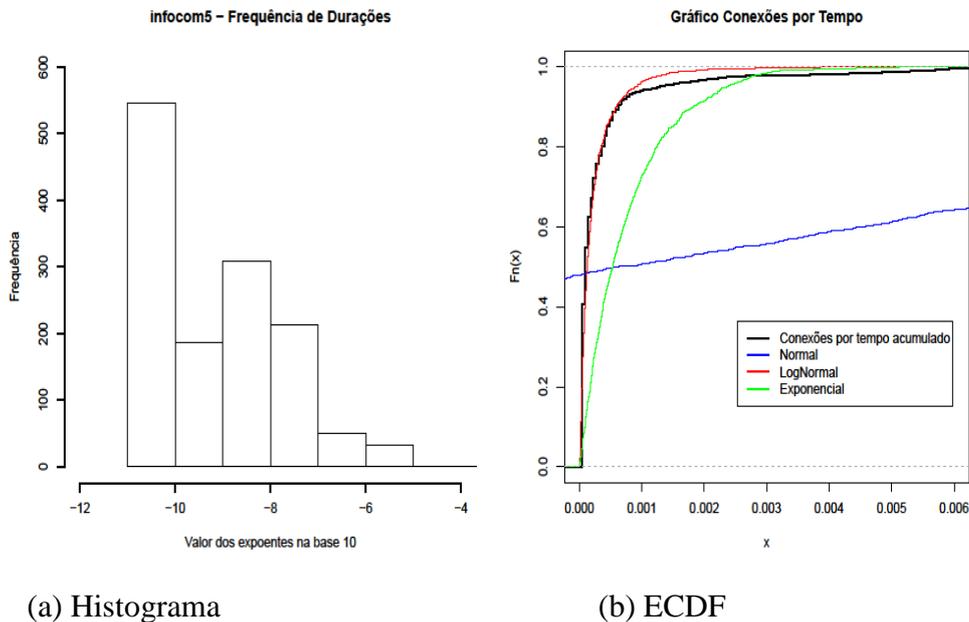


Figura 14: Análise da Frequência de Duração dos Contatos no cenário Infocom5

O histograma, mostrado na Figura 14(a), ilustra o comportamento da variação dos valores calculados da frequência de duração dos contatos a cada instante da coleta dos dados. Os valores do eixo 'x' do gráfico encontram-se em escala logarítmica, representando, cada um, um expoente e todos na base 10. A Figura 14(a) mostra que o ápice de frequência novamente se encontra na parte esquerda do gráfico, mostrando que seus valores são pequenos, da ordem de 10^{-10} . Lembrando que o arquivo que possui os valores das frequências de duração dos contatos foi normalizado, ou seja, todos os valores foram colocados em intervalos entre zero e 1. A média dos valores, nesse caso, se aproxima de 0.0007 e o desvio padrão é igual a, aproximadamente, 0,01.

Dado que a coleta de dados do cenário Infocom5 possui uma quantidade total de 22.459 contatos, e a média dos valores ocorreu, segundo o histograma, aproximadamente 1.400 vezes, conclui-se que o valor médio da métrica da frequência da duração dos contatos nesse cenário ocorreu em, aproximadamente, 6% da quantidade total de nós dele.

A Figura 14(b) ilustra exatamente o mostrado na Tabela 4, ou seja, a curva da distribuição Lognormal sendo a mais próxima da curva ECDF da métrica, no caso, estando sobrepostas praticamente em todo momento no gráfico e a curva da distribuição Exponencial sendo a segunda mais próxima, começando também sobreposta, mas depois apenas estando próxima na maior parte do gráfico.

6.3 Infocom6

Como mencionado anteriormente, esse cenário consiste em uma coleta de dados realizada na Conferência IEEE Infocom 2006, realizada na Espanha. Foram distribuídos 98 dispositivos a estudantes presentes no evento para a realização do tráfego de dados.

A Tabela 6, a seguir, mostra os resultados estatísticos da análise das métricas dentro desse cenário. Novamente, os valores em verde mostram que a distribuição referenciada na coluna é a mais aderente à curva ECDF da métrica relacionada na linha. Note que o teste de aderência entre a curva da distribuição Weibull e a curva ECDF da métrica de frequência da duração de contatos no tempo, novamente, não foi realizado. Isso se deve, novamente, a problemas de propriedade entre a distribuição e a métrica calculada.

| Métricas/Distribuições | Normal | Lognormal | Exponencial | Weibull |
|-------------------------------------|--------------|--------------|--------------|---------------|
| Densidade | 0.000155848 | 0.0002574867 | 0.0002090516 | 0.0001551104 |
| Grau Médio | 0.3665933 | 0.6055193 | 0.4917429 | 0.3637575 |
| Coefficiente de Clusterização Médio | 0.002705003 | 0.003279637 | 0.003646863 | 0.002470236 |
| Frequência do grau de nós no tempo | 0.0002508747 | 0.0001302232 | 0.0001302308 | Não calculado |

Tabela 6: Resultados do MSE das métricas do cenário Infocom6

Como se pode ver na Tabela 6, quanto à densidade, a distribuição Weibull é a mais aderente à sua curva ECDF, tendo a distribuição Normal também próxima, porém com uma diferença bem pequena para mais. As demais distribuições também possuem uma diferença pequena para as duas mais próximas.

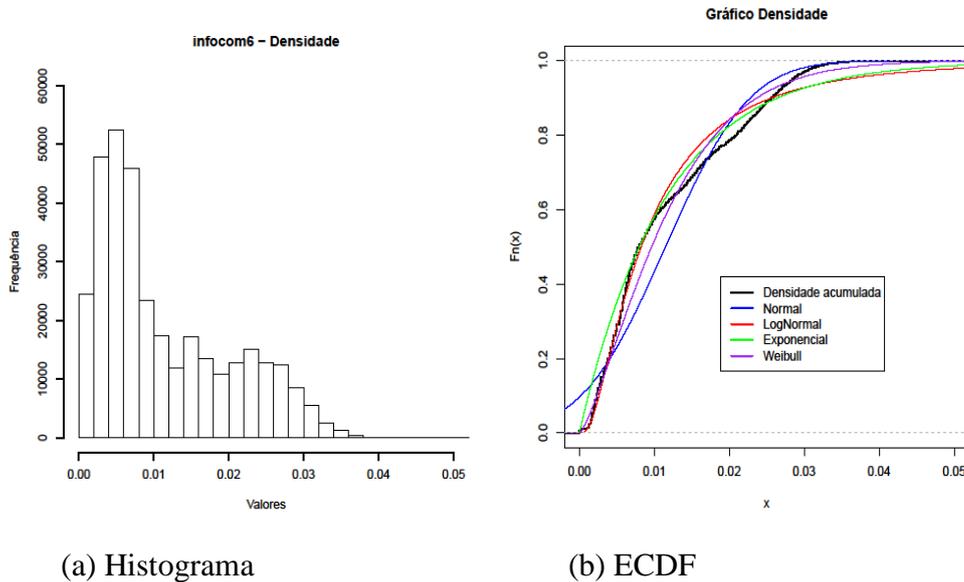


Figura 15: Análise da Densidade no cenário Infocom6

O histograma, mostrado na Figura 15(a), ilustra o comportamento da variação dos valores calculados da densidade a cada instante da coleta dos dados. Ela mostra que o ápice de frequência novamente se encontra na parte esquerda do gráfico, mas com

valores bem distribuídos entre os intervalos, ou seja, muitos instantes com valores cobrindo a maioria dos intervalos. O valor médio da métrica, nesse caso, é de, aproximadamente, 0,01 e o desvio padrão é igual a, aproximadamente, 0,009.

Dado que a coleta de dados do cenário Infocom6 possui uma duração de 328.371 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 51.000 instantes, conclui-se que o valor médio da métrica da densidade nesse cenário ocorreu em, aproximadamente, 16% do tempo total dele.

A Figura 15(b) mostra as quatro curvas das distribuições bem próximas no gráfico, justificando a pequena diferença entre os resultados do MSE das distribuições, na Tabela 6, com relação a essa métrica. Em casos como esse, graficamente não é possível saber qual distribuição é a mais aderente, somente é possível saber utilizando os teste de aderência que, no caso deste trabalho, é o MSE.

Quanto ao grau médio, a Tabela 6 mostra que as distribuições, novamente, possuem pequena diferença em relação a seus resultados de teste de aderência, tendo a distribuição Weibull como a distribuição mais aderente à curva ECDF da métrica, seguida de perto pela distribuição Normal, com uma diferença muito pequena para mais quanto ao resultado do teste, em relação à Weibull.

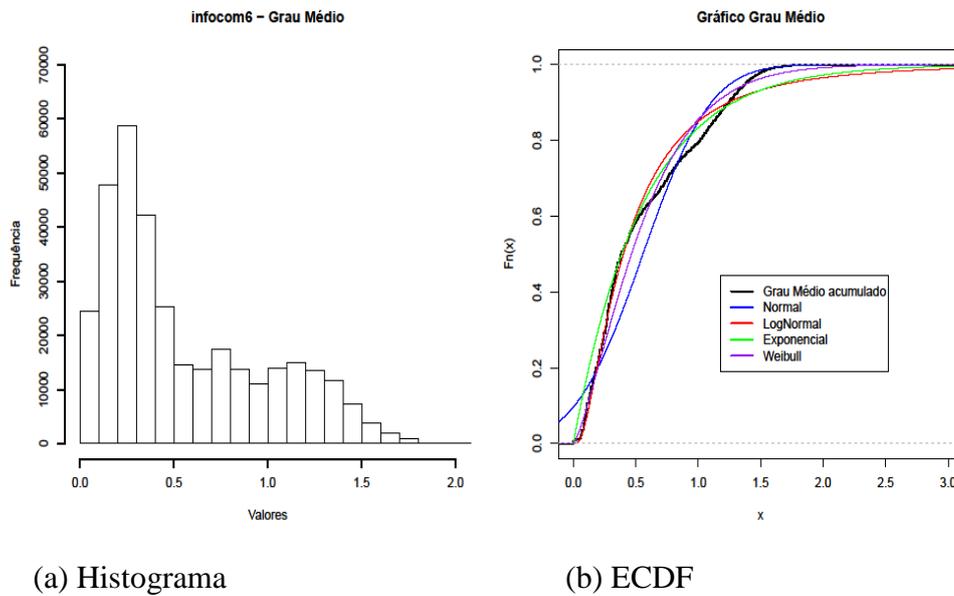


Figura 16: Análise do Grau Médio no cenário Infocom6

O histograma, mostrado na Figura 16(a), ilustra o comportamento da variação dos valores calculados do grau médio a cada instante da coleta dos dados. Ela mostra que o ápice da frequência dos resultados se encontra na parte esquerda do gráfico e que, novamente, os resultados, em geral, abrangem diversos intervalos de valores, obtendo um histograma menos concentrado. A média dos valores, nesse caso, se aproxima de 0,6 e o desvio padrão é igual a, aproximadamente, 0,4.

Dado que a coleta de dados do cenário Infocom6 possui uma duração de 328.371 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 60.000 instantes, conclui-se que o valor médio da métrica do grau médio nesse cenário ocorreu em, aproximadamente, 18% do tempo total dele.

A Figura 16(b) mostra as curvas das quatro distribuições juntas, mais uma vez justificando a pequena diferença de valores entre seus resultados de MSE em relação à curva ECDF do grau médio. Este é mais um caso em que, somente com a análise gráfica, não é possível apontar qual a distribuição mais aderente, apenas com análise estatística, através dos testes de aderência.

Ainda na métrica do grau médio, assim como nos cenários anteriores, foram coletados três nós do *trace* para a análise particular com a curva ECDF da métrica. Os três nós coletados foram: “p11”, por ser o nó com o menor número de contatos; “p24”, por ser o nó que se conectou um número de vezes mais próximo à quantidade média de conexões; e “p87”, por ser o nó de maior número de contatos. A Tabela 7, a seguir, mostra os resultados estatísticos dos testes de aderência entre as curvas ECDF da variação dos graus desses nós no tempo de coleta e a curva ECDF do grau médio deste cenário.

| Nós/ Medida | Grau Médio |
|--------------------|-------------------|
| p11 | 0.8042574 |
| p24 | 2.850363 |
| p87 | 6.077788 |

Tabela 7: Resultados do MSE da variação do grau dos nós no cenário Infocom6

A Tabela 7 mostra que, novamente, o nó de menor ocorrência dentro do *trace* do cenário é o que possui a curva de variação mais aderente à curva ECDF do grau médio, possuindo uma grande diferença com relação aos outros nós quanto ao resultado do teste de aderência com a curva da métrica.

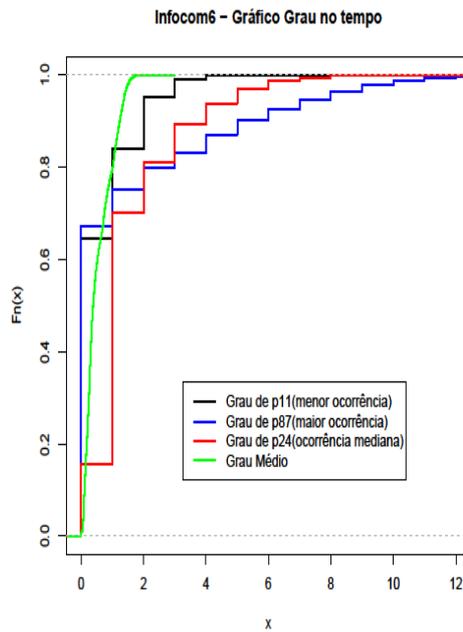


Figura 17: Análise de Variação do Grau dos nós no cenário Infocom6

A Figura 17 mostra a curva referente ao nó de menor ocorrência sendo a mais próxima da curva ECDF referente ao grau médio. No ponto 0,6 e 0,8 do eixo $F_n(x)$, percebe-se que sua curva se sobrepõe à curva ECDF do grau médio, enquanto a curva referente ao nó de ocorrência mediana se sobrepõe apenas uma vez. Nota-se também que há um momento, entre os pontos zero e 0,2 do eixo $F_n(x)$, em que as três curvas referentes aos nós se sobrepõem. Com os resultados das três análises feitas usando o grau médio, conclui-se que, em geral, o resultado do grau médio por instante nesse cenário foi baixo.

Quanto ao coeficiente de agrupamento, a Tabela 6 mostra que os resultados do teste MSE das distribuições possuem, novamente, uma pequena diferença entre si, tendo a distribuição Weibull como a mais aderente à curva ECDF do coeficiente de agrupamento e a distribuição Normal como a segunda mais aderente.

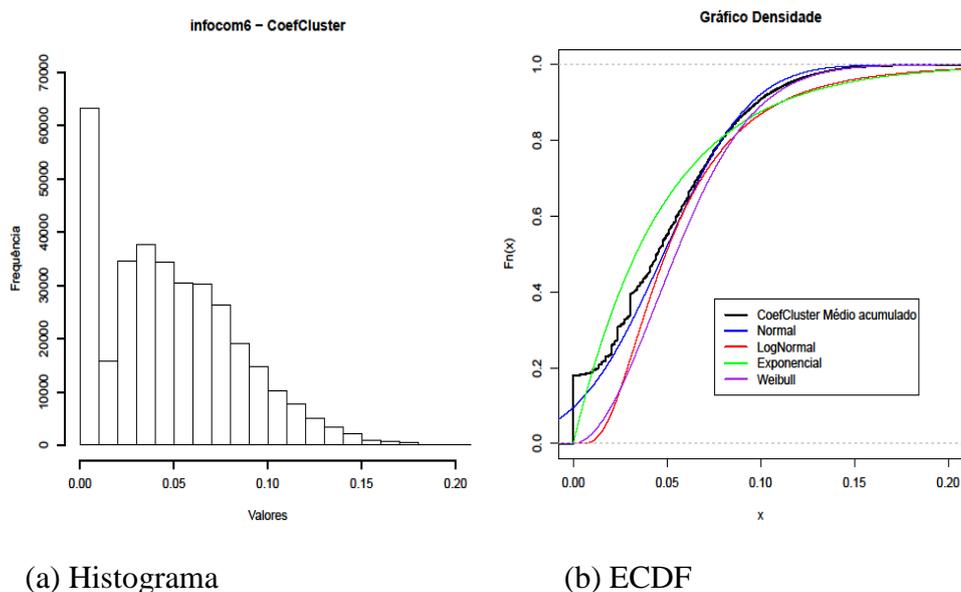


Figura 18: Análise do Coeficiente de Agrupamento no cenário Infocom6

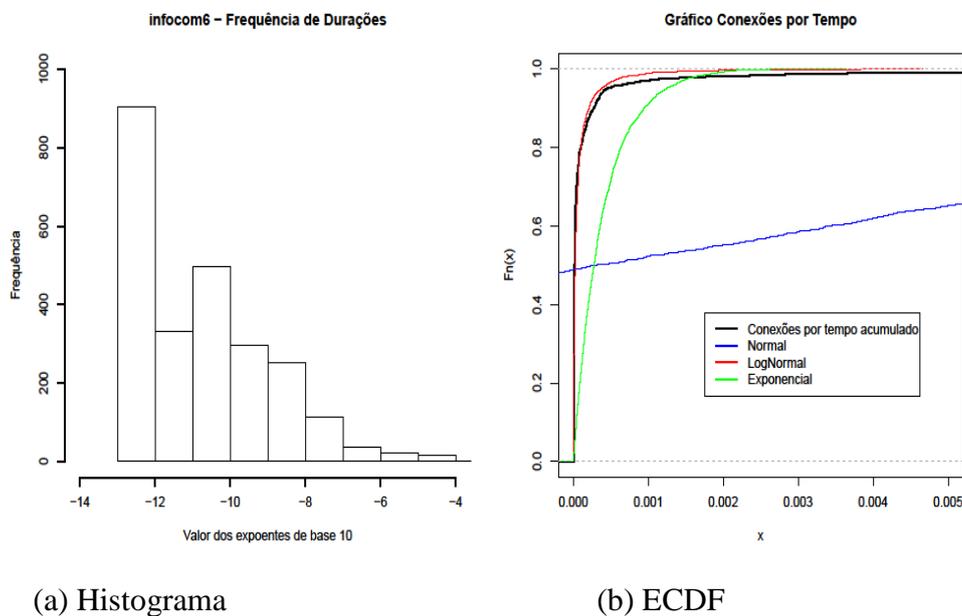
O histograma, mostrado na Figura 18(a), ilustra o comportamento da variação dos valores calculados do coeficiente de agrupamento a cada instante da coleta dos dados. Ela mostra que o ápice da frequência dos resultados se encontra na parte esquerda do gráfico e que, novamente, os resultados, em geral, abrangem diversos intervalos de valores, obtendo um histograma menos concentrado. A média dos valores, nesse caso, se aproxima de 0,05 e o valor do desvio padrão de, aproximadamente, 0,04.

Dado que a coleta de dados do cenário Infocom6 possui uma duração de 328.371 segundos, e a média dos valores ocorreu, segundo o histograma, aproximadamente em 65.000 instantes, conclui-se que o valor médio da métrica do grau médio nesse cenário ocorreu em, aproximadamente, 20% do tempo total dele.

A Figura 18(b) mostra as curvas das quatro distribuições juntas, mais uma vez justificando a pequena diferença de valores entre seus resultados de MSE em relação à curva ECDF do grau médio. Este, também, é um caso em que, somente com a análise gráfica, não é possível apontar qual a distribuição mais aderente, apenas com análise estatística, através dos testes de aderência. Com esse resultado, conclui-se que, no geral,

as conexões nesse cenário foram esparsas, distribuídas, não muito concentradas em nós vizinhos.

Quanto à frequência da duração dos contatos nesse cenário, a Tabela 6 mostra que as distribuições estão próximas com relação ao resultado do teste de aderência, e que a distribuição Lognormal é a mais aderente à curva ECDF da métrica. Como mencionado anteriormente, o resultado do MSE com relação à distribuição Weibull não foi calculado por questões de propriedade da distribuição.



(a) Histograma

(b) ECDF

Figura 19: Análise da Frequência de Duração dos Contatos no cenário Infocom6

O histograma, mostrado na Figura 19(a), ilustra o comportamento da variação dos valores calculados da frequência de duração dos contatos a cada instante da coleta dos dados. Os valores do eixo 'x' desse gráfico se encontram em escala logarítmica, representando, cada um, um expoente, pertencente à base 10. A Figura 19(a) mostra que o ápice de frequência novamente se encontra na parte esquerda do gráfico, mostrando que seus valores, em geral, são pequenos, na ordem de 10^{-12} . Lembrando que o arquivo que possui os valores das frequências de duração dos contatos foi normalizado, ou seja,

todos os valores foram colocados em intervalos entre zero e 1. A média dos valores, nesse caso, se aproxima de 0.0004 e o desvio padrão é igual a, aproximadamente, 0,01.

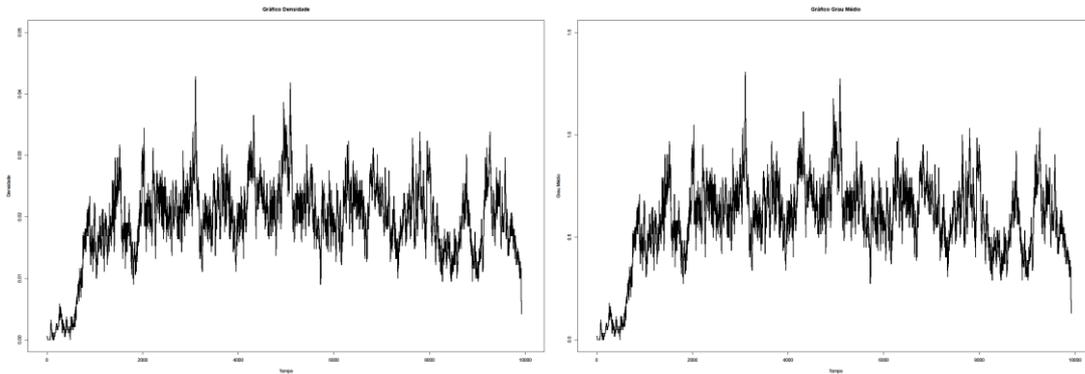
Dado que a coleta de dados do cenário Infocom6 possui uma quantidade total de 65.536 contatos, e a média dos valores ocorreu, segundo o histograma, aproximadamente 2.500 vezes, conclui-se que o valor médio da métrica da frequência da duração dos contatos nesse cenário ocorreu em, aproximadamente, 4% da quantidade total de nós dele.

A Figura 19(b) ilustra exatamente o mostrado na Tabela 6, ou seja, a curva da distribuição Lognormal sendo a mais próxima da curva ECDF da métrica, no caso, estando sobrepostas praticamente em todo momento no gráfico e a curva da distribuição Exponencial sendo a segunda mais próxima, começando também sobreposta, mas depois apenas estando próxima na maior parte do gráfico. Esses resultados mostram que, no geral, as conexões tiveram uma curta duração.

Após a análise de todas as métricas nos três cenários, percebeu-se um certo padrão dentro dos resultados: em, praticamente, todos os casos, as curvas das distribuições estiveram juntas nos gráficos, tendo, em geral, a distribuição Weibull como a mais aderente. Quanto aos histogramas, em todos os casos, o ápice da frequência dos dados encontrou-se na parte esquerda, ou seja, os resultados das métricas, em geral, foram baixos, havendo casos em que houve uma diversidade de intervalos e outros em que foi concentrado em apenas um intervalo.

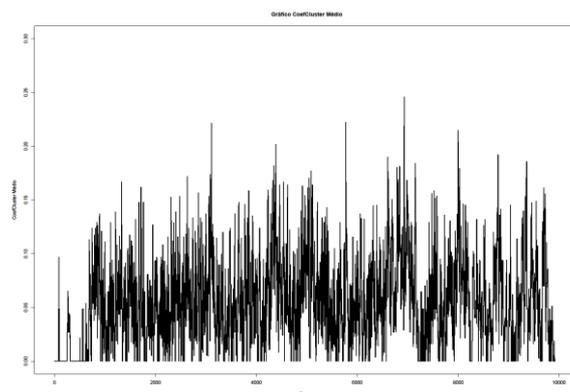
6.4 Análise das Métricas no tempo

Esta seção será destinada à análise dos resultados no tempo das seguintes métricas: densidade, grau médio e coeficiente de agrupamento. A seguir serão mostradas imagens dos gráficos que ilustram o comportamento desses resultados em cada um dos cenários.



a) Densidade

b) Grau Médio



c) Coeficiente de Agrupamento

Figura 20: Análise dos resultados no tempo das métricas no cenário Rollernet

A Figura 20 mostra o comportamento dos resultados das métricas durante o tempo dentro do cenário. Pode-se perceber que as três possuem um comportamento parecido, principalmente a Densidade e o Grau Médio, que são quase idênticos. Esse comportamento parecido entre as três métricas se deve muito à característica do próprio cenário, que se tratou de um evento contínuo, sem interrupções. Essa característica do evento proporcionou, durante todo o tempo, uma proximidade entre as pessoas com os dispositivos, facilitando os contatos, fazendo com que a densidade dos contatos fosse

constantemente alta, que o número médio de contatos (grau médio) fosse também constantemente alto e que a quantidade de contatos entre vizinhos (coeficiente de agrupamento) também fosse constantemente alto, visto que estavam várias pessoas próximas.

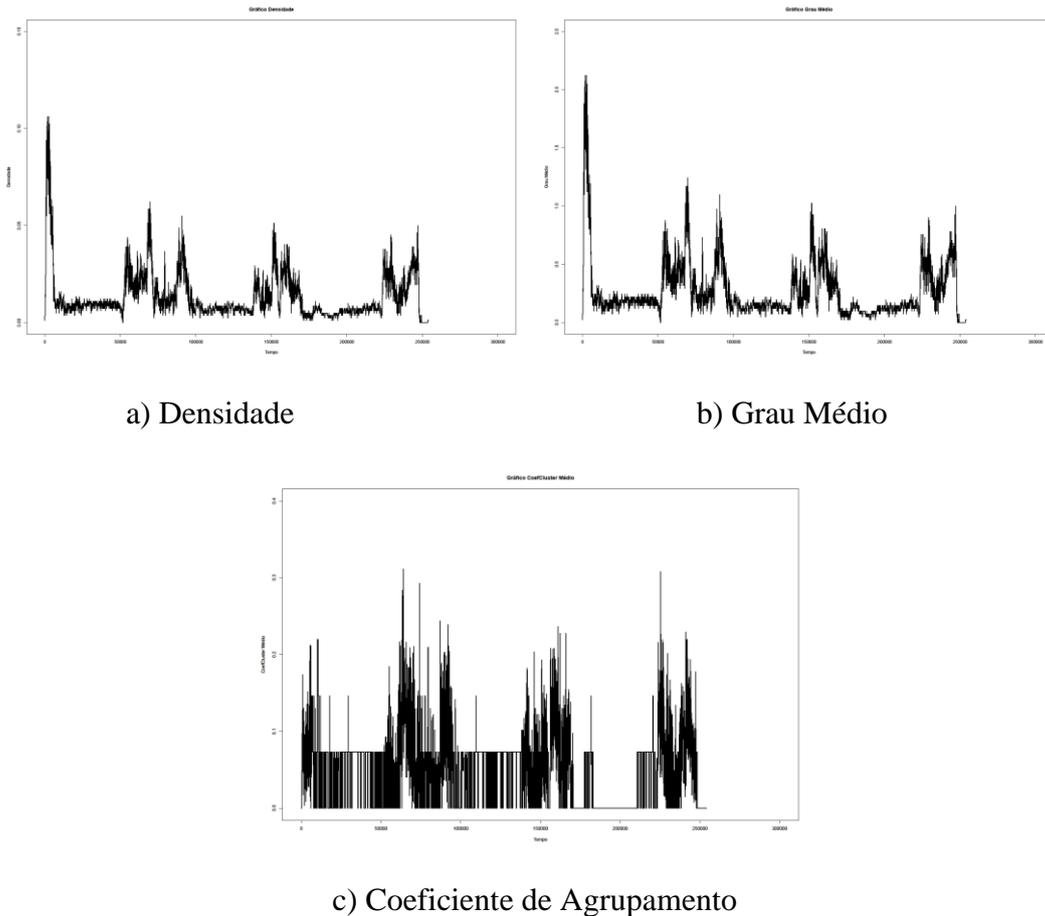


Figura 21: Análise dos resultados no tempo das métricas no cenário Infocom5

A Figura 21 mostra o comportamento dos resultados das métricas durante o tempo no cenário Infocom5. Pode-se notar que as três medidas possuem um comportamento parecido entre si, mas diferente em relação ao cenário do Rollernet, analisado anteriormente. Essa diferença se dá pelo estilo do evento onde os contatos foram realizados. Enquanto o cenário do Rollernet foi um evento contínuo de curta duração, os contatos desse cenário Infocom5 foram realizados em uma conferência de 4 dias. E essa conferência não durou continuamente os 4 dias, sendo um evento com interrupções. Essas interrupções explicam os picos registrados pelos gráficos, onde, cada um, tem 4

picos, registrados durante o tempo em que a conferência foi realizada nos 4 dias, quando as pessoas com os dispositivos estavam próximas e estabelecendo contatos. Os espaços temporais onde os valores são baixos são resultantes dos instantes onde não houve a conferência e as pessoas estavam separadas, não estabelecendo contatos.

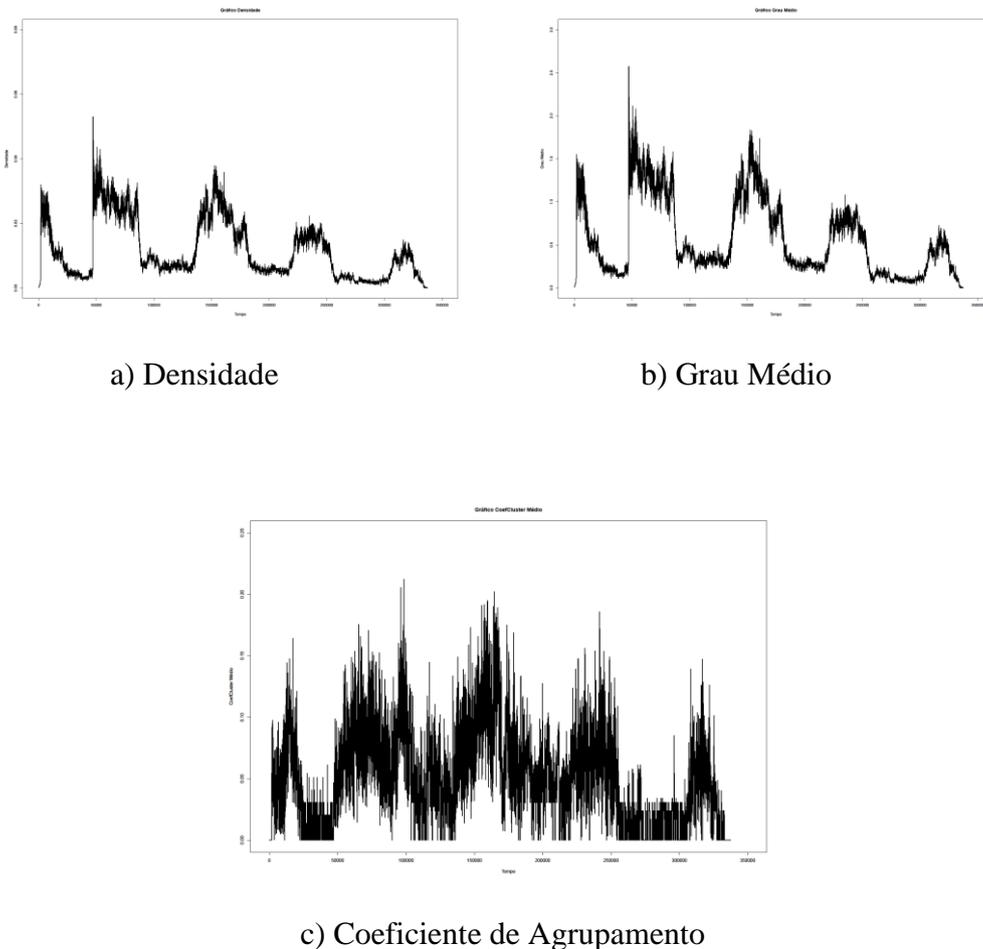


Figura 22: Análise dos resultados no tempo das métricas no cenário Infocom6

A Figura 22 mostra o comportamento dos resultados das métricas durante o tempo no cenário Infocom6. Pode-se notar que o comportamento das métricas nesse cenário se assemelha muito com o comportamento das métricas no cenário Infocom5, anteriormente analisado. Essa semelhança é explicada pelo fato de que os eventos possuem características idênticas, sendo, no fundo, o mesmo evento, apenas um (Infocom6) sendo uma edição posterior ao outro (Infocom5). Porém, pode-se perceber que esse cenário teve uma pequena diferença com relação ao Infocom5: ele possui mais

picos que o Infocom5, o que significa que, durante os também 4 dias de evento, houve mais tempo de pessoas próximas e realizando contatos.

7. Conclusão

Ao longo deste trabalho, foi realizado um estudo sobre o comportamento dos dados em três cenários reais de mobilidade humana, usando arquitetura de redes DTN, através da geração de grafos de conectividade e implementação de métricas. Através da análise dessas métricas, foi possível obter conclusões sobre o comportamento desses dados. Neste capítulo serão detalhados os passos e conclusões tiradas desse trabalho, além das dificuldades encontradas.

7.1 Considerações Finais

Como mencionado no Capítulo 4, foram escolhidos três cenários reais de mobilidade humana do repositório de dados CRAWDAD. Neste repositório, foi coletado um arquivo contendo todas as conexões durante o tempo do experimento realizado, com os nós participantes e o momento e duração em que ocorreu cada conexão. Esse arquivo é chamado de *trace* do cenário.

O próximo passo foi, como mencionado no Capítulo 5 (Seção 1), a implementação de uma rotina que convertesse esse *trace* para um padrão que facilitasse a implementação das métricas temporais para a análise de cada cenário. A partir disso, foram escolhidas as métricas, mencionadas e definidas na Seção 2, e as mesmas foram implementadas baseadas no padrão convertido dos *traces*. A partir do momento em que as métricas terminaram de ser implementadas, seus resultados foram gerados.

Com os resultados gerados, o processo de análise de comportamento começou, como mencionado do Capítulo 6. Foi utilizado o *software* R para fazer análises gráficas e estatísticas dos dados. Primeiramente, foi gerado, para cada resultado de cada métrica, de cada cenário, um histograma, possibilitando a visualização da frequência dos dados, ao longo do tempo, contidos no arquivo de resultados.

Posteriormente, foi gerado o gráfico ECDF de cada resultado, de cada métrica. Este gráfico mostra a curva cumulativa de cada métrica ao longo do tempo. Com os gráficos ECDF gerados, foram escolhidas quatro distribuições de probabilidade (Normal, Lognormal, Exponencial e Weibull) para analisar o comportamento de cada curva ECDF de cada métrica. O intuito da análise é descobrir a qual distribuição de probabilidade cada curva ECDF de cada métrica tinha o comportamento mais parecido. Para descobrir isso, foi utilizado um teste de aderência, o MSE, que consiste em calcular a diferença de cada ponto entre a curva ECDF e a curva da distribuição, e, ao final, gerar um resultado médio dessa diferença e, o que tiver o menor resultado, será a distribuição mais aderente.

A Tabela 8, a seguir, mostra o resultado dessa análise para cada métrica de cada cenário.

| Métricas/Cenários | Rollernet | Infocom5 | Infocom6 |
|--|------------------|-----------------|-----------------|
| Densidade | Weibull | Weibull | Weibull |
| Grau Médio | Weibull | Weibull | Weibull |
| Coefficiente de Clusterização Médio | Weibull | Exponencial | Weibull |
| Frequência da duração dos contatos | Exponencial | Lognormal | Lognormal |

Tabela 8: Resultados da Análise do comportamento das curvas ECDF

Como pode ser visto na Tabela 8, a maioria das métricas dos cenários possuem curvas ECDF com comportamento mais semelhante ao da distribuição Weibull, havendo apenas quatro métricas que não possuem o comportamento parecido com essa distribuição (lembrando que, em duas delas, nas frequências dos cenários Infocom5 e Infocom6, a distribuição Weibull não pôde ser gerada).

Observa-se que, com relação às métricas da densidade e grau médio, em todos os cenários suas curvas ECDF foram mais aderentes à distribuição Weibull. Quanto ao coeficiente de agrupamento, apenas no cenário Infocom5, sua curva não teve comportamento mais aderente ao da distribuição Weibull, sendo mais aderente à distribuição Exponencial. Quanto à frequência de duração dos contatos, sua curva teve comportamento mais aderente à distribuição Lognormal nos cenários Infocom5 e Infocom6, enquanto, no Rollernet, foi mais aderente à distribuição Exponencial. Percebeu-se que nenhuma das métricas foi mais aderente à distribuição Normal em nenhum dos cenários.

Através da leitura dos trabalhos relacionados, foi possível identificar um esforço para elaboração de novos algoritmos e protocolos que permitissem um uso mais otimizado da tecnologia DTN. Para futuros trabalhos, o conhecimento gerado pela presente análise de mobilidade poderá servir de apoio para criação de uma ferramenta que agilize a distribuição de informação na rede, bem como auxilie na escolha de um melhor caminho para que os bundles cheguem ao seu destino. Tal desenvolvimento poderá tirar vantagem, por exemplo, do conhecimento de que, em todos os cenários aqui analisados, a densidade temporal da rede tem comportamento ajustado ao de uma função Weibull.

7.2 Dificuldades encontradas ao longo deste trabalho

Com relação às dificuldades encontradas, basicamente houve alguma em cada passo, porém, serão descritas as maiores: na fase de implementação das métricas e análise dos resultados.

Na fase de implementação das métricas, as dificuldades foram encontradas na definição de quais métricas seriam mais relevantes para o cenário, o que possibilitaria gerar conclusões mais significantes. Para isso, foi necessária uma busca sobre as possíveis métricas e o que nos agregariam de informação sobre os cenários. A partir daí, a única dificuldade, nessa fase do trabalho, foi a implementação das métricas.

Na fase da análise, as dificuldades foram encontradas no uso do *software* R, com seus comandos próprios, na definição dos tipos de gráficos a serem gerados e o teste de aderência a ser utilizado. Outra dificuldade encontrada foi a geração das curvas das distribuições de probabilidade no mesmo gráfico da curva ECDF das métricas.

8. Referências Bibliográficas

- [1] KEVIN FALL e STEPHEN FARRELL. DTN: An Architectural Retrospective. IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 26, NO. 5, JUNE 2008.
- [2] E. YONEKI, P. HUI, S. CHAN, and J.CROWCROFT. A socio-aware overlay for publish/subscribe communication in delay tolerant networks. In Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems, MSWiM '07, 225-234. ACM, New York, NY, USA (2007).
- [3] FABRÍCIO BENEVENUTO, JUSSARA M. ALMEIDA e ALTIGRAN S. SILVA. Explorando Redes Sociais Online: Da Coleta e Análise de Grandes Bases de Dados às Aplicações. Mini-curso SBRC (2011).
- [4] ROBERT HILBRICH. Social Network Analysis using Graph Metrics of Web-based Social Networks. SNA Presentation, Department of Computer Science Humboldt University at Berlin. November 27, 2007.
- [5] THEUS HOSSMANN, THRASYVOULOS SPYROPOULOS, FRANCK LEGENDRE. Putting Contacts into Context: Mobility Modeling beyond Inter-Contact Times. NETSCICOM (2011).
- [6] JOHN TANG, MIRCO MUSOLESI, CECILIA MASCOLO, VITO LATORA. Temporal Distance Metrics for Social Network Analysis. SIGCOMM, (2009).
- [7] HEMAL SHAH, YOGESHWAR KOSTA, VIKRANT PATE. Characterizing and Evaluation :Temporal properties of real and synthetic datasets for DTN. Ganpat University, Marwadi Education Foundation.

- [8] RAFAEL LOPES BEZERRA, Análise da Conectividade em Redes Móveis utilizando dados obtidos da mobilidade humana. Universidade Federal do Rio de Janeiro (UFRJ), Março/2009.
- [9] V. LENDERS, G. KARLSSON, and M. MAY. Wireless ad hoc podcasting. 2007 4th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks, 273-283 (2007).
- [10] V. LENDERS, M. MAY, G. KARLSSON, and C. WACHA. Wireless ad hoc podcasting. SIGMOBILE Mob. Comput. Commun. Rev., 12, 65-67 (2008).
- [11] J. GREIFENBERG, D. KUTSCHER. Efficient publish/subscribe-based multicast for opportunistic networking with self-organized resource utilization. In Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops, 1708-1714. IEEE Computer Society, Washington, DC, USA (2008).
- [12] C. BOLDRINI, M. CONTI, and A. PASSARELLA. Design and performance evaluation of ContentPlace, a socialaware data dissemination system for opportunistic networks. Computer Networks, 54, 589-604 (2010).
- [13] KEVIN ALMEROOTH, GUNNAR KARLSSON, CECILIA MASCOLO and JORG OTT. Impact of Human Mobility on Communications: Measurement, Analysis, Modeling, and Simulation. Dagstuhl Seminar. October, 2010.
- [14] BERK BIRAND, MURTAZA ZAFERY, GIL ZUSSMAN and KANG-WON LEE. Dynamic Graph Properties of Mobile Networks under Levy Walk Mobility. Department of Electrical Engineering, Columbia University, New York, NY.
- [15] PAN HUI, JON CROWCROFT, EIKO YONEK. BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks. Cambridge University (2008).

[16] PADMA MUNDUR, MATTHEW SELIGMAN. Delay Tolerant Network Routing: Beyond Epidemic Routing. College Park, ISWPC 2008.

[17] THRASYVOULOS SPYROPOULOS, KONSTANTINOS PSOUNIS and CAULIGI S. RAGHAVENDRA. Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks. Department of Electrical Engineering, USC (2005).

[18] ROBERT HANNEMAN and MARK RIDDLE. Introduction to social network methods. University of California, Riverside. (2005).

[19] CARLOS ALBERTO VIEIRA CAMPOS. Caracterização da Mobilidade Humana e de Modelos de Mobilidade e suas Influências nas Redes Móveis. COPPE/UFRJ. (Setembro/2009).

[20] AUGUSTIN CHAINTREAU, PAN HUI, JON CROWCROFT, FELLOW, IEEE, CHRISTOPHE DIOT, RICHARD GASS, and JAMES SCOTT. Impact of Human Mobility on Opportunistic Forwarding Algorithms. (2007)

[21] AUGUSTIN CHAINTREAU, ABDERRAHMEN MTIBAA, LAURENT MASSOULIE, CHRISTOPHE DIOT. The Diameter of Opportunistic Mobile Networks. CoNEXT'07, December 10-13, 2007, New York