

# GROUPS-NET: Roteamento Ciente de Encontros de Grupos em Redes Móveis D2D

Ivan Oliveira Nunes<sup>1</sup>, Pedro O. S. Vaz de Melo<sup>1</sup>, Antonio A. F. Loureiro<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)  
Av. Antônio Carlos, 6627 – Pampulha, 31270-901- Belo Horizonte - MG - Brazil

{ivanolive,olmo,loureiro}@dcc.ufmg.br

**Abstract.** *The increasing ubiquity of mobile devices with inter-connection capability leveraged the possibility of device-to-device communication. In the next generation cellular networks, device-to-device communication is already considered a fundamental feature. Outside cellular communication context, such kind of mobile Ad-hoc networks are also referred as Delay Tolerant Networks (DTNs). A fundamental problem of DTNs is on how to define forwarding algorithms that achieve, at the same time, high delivery ratio and low network overhead. In the past decade, several algorithms were proposed to achieve those goals. Among them, probabilistic and social aware approaches have remarked themselves as the most successful strategies. However, existent social aware algorithms do not account for the underneath dynamics and evolving nature of social groups. In this work, we propose a probabilistic opportunistic routing protocol which is aware of group meetings and how these meeting properties evolve over time. In large scale scenarios, our algorithm achieve approximately the same delivery ratio of state-of-art solution with up to 40% less network overhead.*

**Resumo.** *A crescente ubiquidade de dispositivos móveis com capacidade de interconexão alavancou a possibilidade de comunicação dispositivo-dispositivo. Na próxima geração de comunicação celular, a comunicação dispositivo-dispositivo já é considerada uma funcionalidade fundamental. Fora do contexto de redes celulares, esse tipo de rede Ad-hoc também é referida como rede tolerante a atrasos (do inglês, DTN). Um problema fundamental em DTNs está em como definir algoritmos de encaminhamento de mensagens que, ao mesmo tempo, alcancem alta taxa de entrega aos destinatários e baixa sobrecarga na rede. Nas últimas décadas vários algoritmos foram propostos. Entre eles, algoritmos probabilísticos e algoritmos cientes de contexto social se destacaram como estratégias de maior sucesso. Porém, os algoritmos cientes de contexto social existentes na literatura não consideram a natureza dinâmica e evolutiva das relações de grupos sociais. Neste trabalho, propõe-se um protocolo probabilístico de roteamento oportunístico que considera reuniões de grupos e como as propriedades dessas reuniões evoluem ao longo do tempo. Para testes realizados em cenários de larga escala, o algoritmo proposto apresentou, em relação à solução do estado da arte, aproximadamente a mesma taxa de entrega com redução de 40% na sobrecarga da rede.*

## 1. Introdução

A crescente demanda por conteúdos populares com grande volume de dados como vídeos, imagens e músicas alavancou a utilização de comunicação dispositivo-dispositivo (D2D)

na próxima geração das redes celulares [Li et al. 2014]. Em redes celulares com suporte a D2D, os dispositivos podem se comunicar diretamente sob um moderado controle das estações bases. Essa nova arquitetura permite que algoritmos de controle centralizados na estação base atuem dinamicamente na política de encaminhamento de dados dos dispositivos. Além disso, muitas aplicações sensíveis ao contexto, que envolvem a descoberta e comunicação com dispositivos próximos, podem se beneficiar da comunicação D2D, economizando em potência de transmissão (já que os dispositivos estarão fisicamente próximos) e aumentando a eficiência espectral da rede celular [Fodor et al. 2012]. Espera-se que a comunicação D2D seja uma funcionalidade chave na próxima geração de comunicação celular [Lei et al. 2012].

Nesse contexto, a utilização de algoritmos bem estabelecidos destinados a Redes Tolerantes a Atraso (DTNs) é vista como uma forma de viabilizar a comunicação D2D em redes celulares. Esses algoritmos de encaminhamento em redes DTN, por sua vez, baseiam-se primordialmente na estrutura estática das comunidades sociais que são formadas entre as pessoas para, assim, definir as suas políticas de encaminhamento. Porém, a detecção distribuída de comunidades é uma tarefa complexa. Como discutido por Hui et al. [Hui et al. 2007], algoritmos existentes para detecção distribuída de comunidades atingem, no máximo, 85% de precisão em relação a suas versões centralizadas. Isso claramente impõe dificuldades e vieses à utilização de algoritmos baseados em comunidades. Além disso, algoritmos baseados em comunidades estáticas não consideram a dinamicidade da mobilidade humana, não levando em conta que essas comunidades podem evoluir e se modificar ao longo do tempo. Por fim, a detecção de comunidades depende da calibração de parâmetros para cada cenário específico, o que torna sua utilização inviável em aplicações de tempo real.

Considerando essas questões, no presente trabalho propõe-se um algoritmo de encaminhamento ciente de contexto social que substitui a utilização de comunidades sociais estáticas pelo conhecimento da dinâmica de encontros de grupos. No contexto deste trabalho, um grupo é definido com um conjunto de pessoas que se encontram de forma recorrente de modo que a maioria dessas pessoas esteja sempre presente nos encontros. Ao contrário das comunidades, do ponto de vista de implementação distribuída, é muito simples identificar encontros de grupos. Basta que cada dispositivo armazene a lista de nós que permaneceram próximos a ele por tempo suficiente para considerar que esteja ocorrendo uma interação social. Dessa forma, propõe-se utilizar a ciência de reuniões de grupos em um esquema de roteamento probabilístico grupo-a-grupo que leva uma mensagem de um usuário origem a um usuário destino. As principais contribuições deste trabalho são:

- Uma metodologia para identificar e rastrear encontros de grupos em *traces* de contatos.
- Um algoritmo de roteamento oportunístico para redes D2D sensível ao histórico de encontros sociais de grupos que elimina a necessidade de detecção distribuída de comunidades.
- Uma análise comparativa que avalia a taxa de entrega e sobrecarga da rede em cenários reais (emulados). O algoritmo proposto revela ser mais apropriado que o atual estado da arte em diversas situações.

As demais seções do artigo estão organizadas como a seguir. A Seção 2 revisa alguns trabalhos relacionados e discute as respectivas contribuições nas áreas de mobilidade humana, detecção de comunidades e roteamento oportunístico. A Seção 3 descreve

as etapas metodológicas para identificar e rastrear grupos a partir de *traces* de contatos. Na Seção 4 é introduzido o algoritmo proposto: GROUPS-NET (*Groups ROUTing in Pocket Switched-NETworks*). Na Seção 5 é apresentada a metodologia de testes e a comparação entre os algoritmos GROUPS-NET e Bubble Rap. Finalmente a Seção 6 traz as considerações finais e os trabalhos futuros.

## 2. Trabalhos Relacionados

**Mobilidade Humana.** Soluções de roteamento para ambientes móveis não podem ignorar as propriedades que governam a mobilidade humana. Gonzalez et al. [Gonzalez et al. 2008] utilizaram uma grande base de dados de telefonia, *Call Detail Records* (CDR), para caracterizar padrões individuais de mobilidade. Seus resultados mostram que o deslocamento radial nas trajetórias humanas, denominado *Radius of Gyration*, segue uma distribuição de probabilidades *power-law* com corte exponencial. Suas análises também revelaram que as trajetórias humanas apresentam periodicidade. Em relação às dinâmicas de mobilidade em grandes cidades, alguns trabalhos avaliaram o impacto gerado por eventos de larga escala. Calabrese et al. [Calabrese et al. 2010] analisaram mais de 1 milhão de dados de telefonia celular para revelar o impacto desse tipo de evento no trânsito de veículos e na mobilidade. Segundo os autores, a compreensão da origem das pessoas que comparecem a um dado evento é fundamental para propor políticas para mitigação de congestionamentos no tráfego. Silva et al. [Silva et al. 2014] propuseram uma técnica de visualização, denominada *City Image*, que captura as transições típicas entre pontos de interesse (PoIs) em cidades usando dados públicos coletados da rede social *Foursquare*. “*City Images*” podem ser compreendidas como intuições para os padrões de mobilidade de cada cidade. Existem também diversos trabalhos no sentido de caracterizar e prever condições do trânsito usando diversos tipos de fontes de dados como em [Lu and Cao 2003, Bauza et al. 2010]. Esses estudos investigam, por exemplo, como condições climáticas afetam as condições do tráfego. Nenhum dos trabalhos supracitados, porém, analisou os padrões de mobilidade de grupos.

**Detecção de Comunidades.** Em uma implementação real, a detecção distribuída de grupos é uma tarefa simples, pois um grupo é definido somente pelas pessoas que estão geograficamente próximas durante um período de tempo. No entanto, a tarefa de detectar grupos de forma centralizada em *traces* de contato de larga escala não é simples. Tal tarefa consiste em analisar milhões de contatos que ocorreram das mais diversas formas e durações. Assim, neste trabalho aplicamos a consolidada ideia de detecção de comunidades em grafos para detectar grupos em *traces* de mobilidade. Desde sua introdução, algoritmos de detecção de comunidades em grafos tem atraído grande atenção. Algoritmos de detecção de comunidades podem ser classificados de acordo com duas características: com sobreposição  $X$  sem sobreposição e estáticos  $X$  dinâmicos. Dentre vários algoritmos propostos, [Palla et al. 2005] e [Gregory 2010] se destacaram como dois dos mais populares e efetivos para detecção de comunidades em grafos estáticos. Trabalhos como [Nguyen et al. 2011] focam em propôr adaptações e novos algoritmos voltados especificamente para grafos dinâmicos, considerando questões como eficiência computacional. Existem também esforços que buscam mapear a evolução das comunidades detectadas em diferentes tipos de redes complexas [Palla et al. 2007]. No presente estudo, são utilizadas essas metodologias desenvolvidas, especificamente *Clique Percolation Method* (CPM) [Palla et al. 2005], para detectar reuniões de grupos a partir de *traces* de contatos.

**Roteamento Oportunístico.** Dentre as soluções propostas de roteamento oportunístico, destaca-se o roteamento probabilístico e o roteamento sensível ao contexto social. O

roteamento probabilístico em DTNs foi primeiramente introduzido por Lindgren et al. [Lindgren et al. 2003]. A ideia central é que encontros entre pares que ocorreram mais recentemente são mais prováveis de ocorrerem no futuro próximo. O algoritmo PROPHET, que implementa essa ideia central, obteve grande sucesso, mas foi superado por algoritmos que consideram as relações sociais entre os usuários da DTN. Nesse sentido, Hui et al. [Hui et al. 2011] usaram a estrutura de comunidades sociais em redes móveis combinada com popularidade de nós para projetar um protocolo de encaminhamento de mensagens denominado *Bubble Rap*. Até a presente data, *Bubble Rap* é o algoritmo que apresenta os melhores resultados em termos de maximização da taxa de entrega de mensagens e minimização da sobrecarga da rede. A aplicação do *Bubble Rap* a redes celulares D2D é proposta em [Li et al. 2014]. Diferente de [Hui et al. 2011], a solução proposta neste trabalho busca detectar grupos de pessoas que estão de fato reunidas, próximas espacialmente e temporalmente. Em [Hui et al. 2011], os autores constroem um único grafo que agrega todos os contatos entre pares que ocorreram ao longo da duração da base de dados e identificam comunidades de nós que possuem laços sociais mais fortes. Essa abordagem impede de levar em conta as mudanças no comportamento social ao longo do tempo, i.é., o dinamismo social. Além disso, é necessário detectar comunidades de forma distribuída, o que é, ainda, um grande desafio. A metodologia proposta neste trabalho é capaz de capturar a natureza dinâmica dos encontros sociais e não necessita da detecção dinâmica de comunidades. Finalmente, esse trabalho propõe e avalia comparativamente uma nova forma de utilizar a dinâmica dos encontros sociais para projetar um algoritmo de encaminhamento de mensagens em redes D2D.

### 3. A Dinâmica dos Encontros de Grupos

Embora em um cenário real a detecção distribuída de uma reunião de grupo seja muito simples, para permitir a avaliação experimental do algoritmo proposto é preciso detectar reuniões de grupos a partir de *traces* de contatos, que são o tipo de dado tipicamente utilizado para avaliar algoritmos de encaminhamento D2D. Nesta seção são descritos a metodologia utilizada para detectar grupos a partir de *traces* de contatos entre usuários móveis e os principais resultados que motivaram a proposta de um algoritmo de roteamento oportunístico ciente do contexto de grupos.

#### 3.1. Identificação dos Encontros de Grupos

##### 3.1.1. Modelando *Traces* de Contatos com Grafos

Para detectar encontros de grupos propõe-se um modelo de grafo social temporal. Primeiramente, divide-se a base de dados de contatos em janelas de tempo  $tw$  (a duração ideal para  $tw$  é discutida na Seção 3.1.2). Contatos dentro de uma mesma janela de tempo  $tw$  são agregados em um grafo de contatos  $G_c(V, E[tw = i])$ , onde  $V$  é o conjunto de nós da base de dados (i.e., pessoas) e  $E$  é o conjunto de arestas que representam contatos de proximidade entre um par de entidades em  $V$ . Assim, nesse modelo, o processamento do *trace* de contatos resultará em um conjunto de grafos subsequentes, com arestas ponderadas e não direcionadas:  $S = \{G_c(V, E[tw = 0]), G_c(V, E[tw = 1]), \dots, G_c(V, E[tw = n])\}$ . O peso de cada aresta  $(v, w) \in E$  pode ser calculado *i)* pelo número de contatos entre  $v$  e  $w$  durante a fatia de tempo  $tw$  ou *ii)* pela soma da duração dos contatos entre  $v$  e  $w$ , quando essa informação se encontra disponível na base de dados. Neste trabalho foram utilizadas uma base de dados em que a duração dos contatos não se encontra disponível, por isso nesse caso computa-se o peso das arestas de acordo com a opção *i)*. Também foi utilizada uma base de dados onde a duração dos contatos está disponível e nesse tipo de base foi utilizada a opção *ii)*.

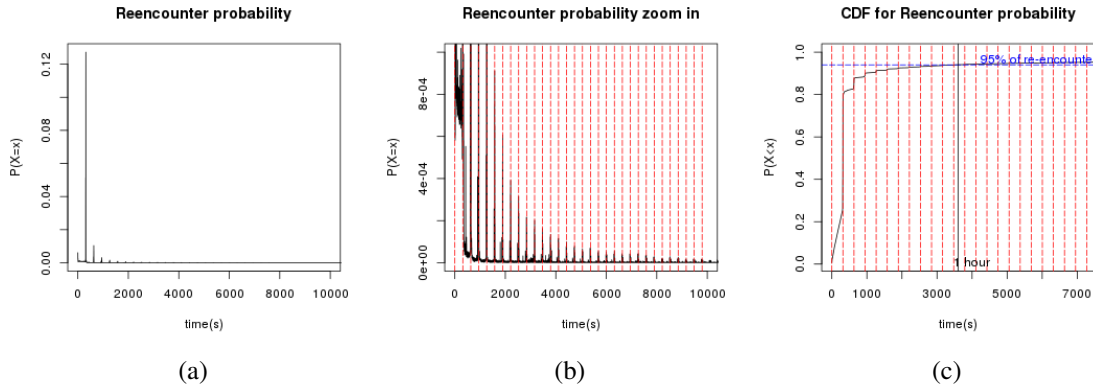


Figura 1. P.D.F. do tempo  $x$  (em segundos) até o próximo contato. Linhas pontilhadas vermelhas mostram um intervalo fixo de 318 segundos em que picos de reencontros ocorrem. Isso mostra que, na maioria das vezes, a aquisição dos dados de proximidade é feita em períodos fixos de 318 segundos. A probabilidade de um par de nós se reencontrar tem distribuição aproximadamente exponencial e 95% dos reencontros ocorrem em menos de uma hora.

### 3.1.2. Caracterização dos Dados

Considerado uma janela de tempo  $tw$  pré-definida, por exemplo 30 minutos, um único contato ao longo de toda essa fatia de tempo não necessariamente significa que este dado par de entidades está interagindo. Isso pode ser causado por uma interseção na trajetória dos nós e não deve necessariamente ser considerada uma interação. Por isso é importante definir um limiar para o peso mínimo  $w_{th}$  de uma aresta para que essa aresta seja considerada representativa de um encontro social. É visível que para definir ambos, o tamanho da janela de tempo  $tw$  e limiar do peso das arestas  $w_{th}$ , é necessário avaliar as propriedades do *trace* de contatos. Nesta seção é realizada uma caracterização das bases de dados utilizadas neste estudo a fim de definir corretamente  $tw$  e  $w_{th}$ . A análise a seguir pode igualmente ser aplicada para outras bases de dados de contatos.

No presente estudo, foram utilizados os traces MIT Reality Mining [Eagle and Pentland 2006] e Dartmouth [Henderson et al. 2008], traces de contatos contendo 80 e 1200 usuários, respectivamente. No trace MIT Reality Mining os usuários monitorados residem em dois prédios da universidade e foram monitorados por diversos meses. Embora o *trace* MIT Reality Mining consista de um cenário específico e de pequena escala, como ele foi o trace utilizado para avaliar o algoritmo do estado da arte Bubble Rap, decidiu-se considerá-lo neste trabalho. Contatos foram registrados quando dois usuários se encontravam a uma distância inferior a 10 metros. Uma entrada no trace de contatos é composta dos IDs do par de nós e da data e hora em que foi registrado o contato do par. Vale mencionar que *traces* de geo-localização (como os obtidos por GPS) podem ser convertidos para *traces* de contatos, definindo uma distância mínima que deva ser considerada um contato entre duas entidades. Por isso, a metodologia aqui apresentada pode também ser aplicada nesse tipo de base de dados. O *trace* Dartmouth monitora todos os usuários de um campus universitário ao longo de dois meses e, além de apresentar os contatos, traz também informação sobre a duração desses contatos. Ao longo dessa seção, devido à limitação de espaço, será apresentada apenas a caracterização para trace MIT Reality Mining, embora a metodologia de caracterização possa ser igualmente aplicada no *trace* Dartmouth e em outros.

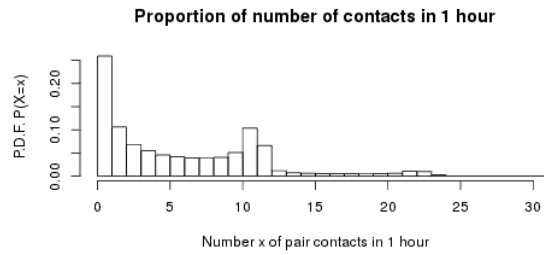


Figura 2. P.D.F. do número de contatos por hora dos pares.

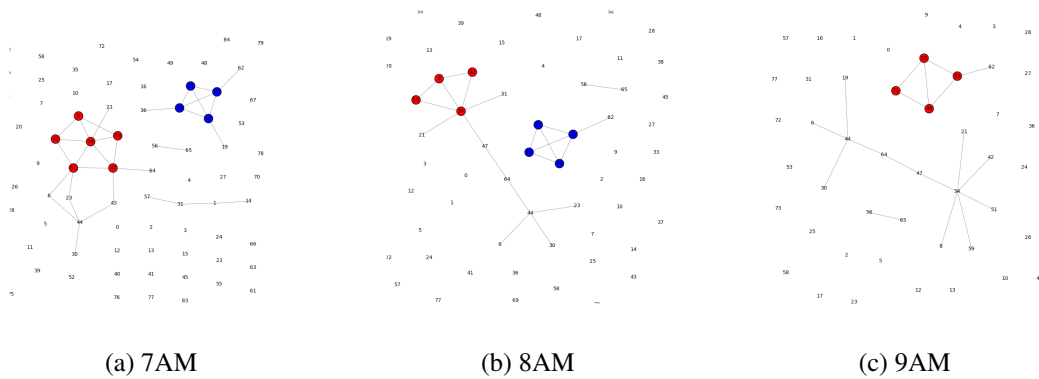


Figura 3. Detecção de grupos usando o CPM, no *trace* MIT Reality Mining, com  $tw = 1h$  em três janelas de tempo consecutivas, em 5 de Fevereiro de 2009. Apenas arestas com  $w_{th} \geq 2$  foram representadas.

Para permitir a detecção e o rastreamento dos grupos é necessário compreender a base de dados utilizada com o objetivo de evitar erros potenciais, tais como erros de amostragem e de inconsistência nos dados, que são gerados no processo de aquisição. Essa avaliação é realizada com o objetivo de definir i) o tamanho de  $tw$ , ou seja, a fatia de tempo ideal para dividir o *trace* e ii)  $w_{th}$ , que é o limiar para o número de contatos (ou duração de contatos) que separa contatos sociais e aleatórios.

Primeiramente, analisa-se o tempo entre os re-encontros de pares, i.e, dado que um par se encontrou, qual é o tempo esperado até o próximo encontro? A Figura 1 mostra que o comportamento dos re-encontros é periódico com moda em períodos de cinco minutos (linhas vermelhas pontilhadas). Esse comportamento indica que o sistema implantado para aquisição dos dados atua de cinco em cinco minutos, na maioria das vezes, mas por alguma razão, pode também atuar em intervalos menores. A partir da função de distribuição acumulada C.D.F. (*cumulative distribution function*) da probabilidade de re-encontros 1-c, nota-se que aproximadamente 95% dos re-encontros podem ser capturados com um  $tw$  de uma hora. Por essa razão, para o *trace* MIT Reality Mining, considerou-se  $tw = 1h$ .

Em seguida, analisa-se a proporção do número de contatos dentro do intervalo  $tw = 1h$  para que seja possível definir  $w_{th}$ . A Figura 2 mostra que 27% dos pares que se encontram em um intervalo de uma hora só se encontram uma vez. Assumimos estes encontros como coincidências nas trajetórias dos pares. Para frequências de encontros entre 2 e 12, o gráfico mostra valores mais uniformes, entre 5 e 10%. Para frequências maiores que 12 encontros, a probabilidade se torna bem baixa, fato que é consistente com

a premissa de que a aquisição dos dados ocorre, na maioria das vezes, em períodos de 5 minutos. De 2 a 12 encontros a função de densidade de probabilidade P.D.F. (probability density function) apresentou valores similares quando comparados com  $P(X = 1)$ . Por essa razão, para o *trace* MIT Reality Mining com  $tw = 1h$ , definiu-se  $w_{th} = 2$ .

Em suma, por meio da caracterização dos dados foi possível definir que dois ou mais contatos no intervalo de uma hora são suficientes para considerar uma interação como social no *trace* MIT Reality Mining. Por meio de análise similar, definiu-se para o *trace* Dartmouth  $tw = 1h$  e  $w_{th} = 10$  minutos de duração de contato.

### 3.1.3. Detecção de Grupos

Após definir valores para  $tw$  e  $w_{th}$ , define-se um grupo como a seguir:

- **Definição de grupo:** Um grupo é um conjunto de nós altamente conectados em  $G_c(V, E[tw = i])$ , que é o grafo gerado a partir da  $i$ -ésima subdivisão do *trace*  $S$ , após eliminar contatos entre pares sem laços sociais e eliminar arestas cujo peso seja inferior ao limiar  $w_{th}$ .

Até aqui, foi estabelecido um modelo para representar interações sociais que consistem de grafos gerados a partir de *traces* de contatos divididos em janelas de tempo. Seguindo a definição acima, faz-se necessário identificar conjuntos de nós *altamente conectados* em um grafo. Para isso, consideramos a vasta literatura de algoritmos de detecção de comunidades em grafos [Fortunato 2010]. Uma comunidade é comumente definida como um conjunto de nós mais densamente interconectados em grafos. Existem na literatura diversos algoritmos para detecção de comunidades em grafos, como [Xu et al. 2013, Nguyen et al. 2011, Gregory 2010]. Dos algoritmos existentes, neste trabalho é utilizado o *Clique Percolation Method* (CPM) [Palla et al. 2005]. As principais razões para se utilizar o CPM são que os membros de suas comunidades podem ser alcançados por conjuntos de nós fortemente conectados e que pode haver sobreposição entre comunidades. Essa última propriedade é essencial, pois a maioria dos grafos sociais caracteriza-se por ter comunidades sobrepostas e aninhadas [Palla et al. 2007]. Para cada grafo de janela de tempo  $G_c(V, E[tw = i])$  computa-se o CPM.

No CPM, uma comunidade é definida como a união de todos os  $k$ -cliques (subgrafos completos de tamanho  $k$ ) que podem alcançar uns aos outros por uma série de  $k$ -cliques adjacentes (onde adjacência significa compartilhar  $k - 1$  nós). O parâmetro  $k$  do CPM limita o tamanho mínimo das comunidades detectadas. Nos últimos anos o CPM consolidou-se como um dos mais bem sucedidos métodos para detecção de comunidades [Peel 2010]. Com o propósito de detectar encontros de grupos, foi atribuído o valor  $k = 3$ , assim, considera-se grupos de três ou mais pessoas. A figura 3 mostra grupos detectados com o CPM no *trace* MIT Reality Mining em 05 de Fevereiro de 2009, às 7, 8 e 9 horas da manhã; janelas de tempo consecutivas de duração de uma hora.

## 3.2. Propriedades dos Encontros de Grupos

Depois de aplicar a metodologia para detecção dos encontros de grupos apresentada na Seção 3, aqui serão discutidas algumas das propriedades mais relevantes identificadas que motivaram a aplicação do conhecimento de grupos no roteamento oportunístico em

D2D/DTNs. Uma visão mais detalhada da dinâmica e da evolução dos encontros de grupos pode ser encontrada em nosso trabalho anterior [Nunes et al. 2016].

A Figura 4-a mostra a frequência dos re-encontros de grupos, i.e., dado que um grupo se encontrou pela primeira vez no tempo  $t = 0$ , como os re-encontros desse grupo estão distribuídos ao longo das horas seguintes? O resultado mostra que a massa de probabilidade se concentra em torno de picos em períodos de 24 horas (representados pelas linhas vermelhas pontilhadas). Isso indica que os encontros de grupos são altamente periódicos considerando periodicidade diária. Também é possível notar que em períodos de 7 dias (representados por linhas pontilhadas verdes) os picos se tornam maiores, evidenciando a presença moderada de periodicidade semanal. Esse resultado faz sentido já que pessoas estão sujeitas horários e rotinas. Esse resultado motivou o experimento seguinte que tenta responder a seguinte pergunta: é possível utilizar encontros de grupos no passado para prever futuros encontros?

Neste experimento, seleciona-se um dos nós do *trace* como origem de uma mensagem e emula-se a transmissão da mensagem de forma epidêmica, ou seja, toda vez que um nó que já recebeu a mensagem encontra com um nó que ainda não a recebeu, a mensagem é propagada. A propagação da mensagem é emulada tendo cada um dos nós do *trace* como origem. Para avaliar a utilidade da informação sobre os contatos de grupos, os nós são então divididos em duas classes: nós que estiveram em grupo com a origem nos últimos 30 dias e nós que não estiveram em grupo com a origem no mesmo período. Atribuiu-se um tempo máximo de 7 dias para propagação epidêmica das mensagens. Ao fim dos sete dias calcula-se a proporção dos nós de cada uma das duas classes que recebeu a mensagem, ou seja, a taxa de entrega.

Como apresentado na figura 4-b, para difusões iniciadas em diferentes meses (a-h) no *trace* MIT Reality Mining, a taxa de entrega para nós que estiveram em grupo com a origem (90%) é significativamente mais alta do que para nós que não estiveram em grupo com a origem (40%). Este resultado motivou a proposta de um algoritmo que explore grupos detectados para realizar roteamento oportunístico em DTNs. Pode-se notar que o mês de junho apresentou taxas de entrega bem menores em relação aos meses anteriores. Isso ocorreu porque o mês de junho é o início do período de férias de verão, quando muitos estudantes deixam o campus.

#### 4. Roteamento Probabilístico Ciente de Grupos

No algoritmo proposto, procura-se explorar o histórico recente de encontros de grupos para definir a rota grupo-a-grupo mais provável para a propagação da mensagem. O algoritmo tem como entrada o conjunto de grupos identificados no período  $T$  (e.g. 24h, 7 dias etc) do passado recente que deve ser considerado para estimar as rotas mais prováveis. Além disso, o parâmetro  $NRP$  indica o número de caminhos redundantes a considerar para o encaminhamento da mensagem.

Para cada grupo identificado no período de  $T$  horas antes do início do roteamento, é criado um nó correspondente em um grafo de rotas. Atribui-se então um peso para o nó, que corresponde ao número de vezes que aquele dado grupo se reuniu nas últimas  $T$  horas. O peso de cada nó  $V_i$  deve ser então normalizado pelo nó de maior peso:  $W(V_i) = W(V_i)/max_i(W(V_i))$ . A premissa é que grupos que se encontram com grande frequência tem maior probabilidade de se reunir novamente no futuro próximo e, dessa forma, o grupo com a maior frequência de encontros terá associado o valor 1. Entre cada par de nós, são criadas arestas de peso  $W(E_{i,j}) = \#(V_i \cap V_j)/max(\#V_i, \#V_j)$ , onde  $\#(V_i \cap$



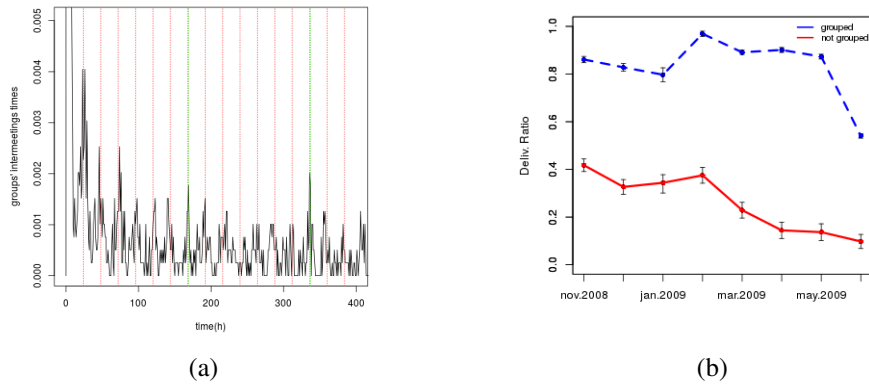


Figura 4. **a)** Probabilidade de um grupo se reencontrar  $t$  horas após ter se encontrado pela primeira vez em  $t = 0$ . Linhas vermelhas representam períodos de 24 horas e linhas verdes períodos de 7 dias. **b)** Taxa de entrega média, para diferentes nós de origem, de Novembro de 2008 a Junho de 2009.

$V_j$ ) é o número de pessoas comuns aos dois grupos e  $\max(\#V_i, \#V_j)$  é o número de componentes do maior dos dois grupos. Com isso, grupos com mais membros em comum passam a ter arestas de maior peso entre si. O peso das arestas está relacionado com a probabilidade de uma mensagem ser transmitida entre o par de grupos. Ao fim dessa inicialização obtém-se um grafo  $G(V, E)$  em que  $V$  é o conjunto dos grupos encontrados e  $E$  é o conjunto de arestas que representam as interseções entre os grupos.

Com o objetivo de encontrar o caminho mais provável em  $G(V, E)$ , recalcula-se o peso de cada aresta  $E_{i,j}$  como  $W(E_{i,j}) = W(V_i) * W(V_j) * W(E_{i,j})$ , onde  $V_i$  e  $V_j$  são os nós da aresta  $E_{i,j}$ . Com isso, considera-se a chance de ambos os grupos ligados pela aresta se reunirem e a chance da mensagem ser levada de um deles para o outro. O novo peso das arestas pode ser entendido como uma aproximação da probabilidade de uma mensagem ser propagada entre os membros do par de grupos no futuro próximo, considerando a probabilidade de reunião de cada um dos grupos e a possibilidade da mensagem ser levada de um grupo para o outro por algum usuário que é membro de ambos. Por fim, para computar as rotas mais prováveis de propagação da mensagem, assume-se independência entre as ocorrências das arestas e a probabilidade de uma rota  $R$  passa a ser calculada de acordo com a Eq. 1.

$$P(R) = \prod W(E_{i,j}), t.q. E_{i,j} \in R \quad (1)$$

Tirando proveito da propriedade de máxima verossimilhança de logaritmos pode-se obter o conjunto de arestas que maximizam a Eq. 1 de acordo com a Eq. 2.

$$\arg.\max(\prod W(E_{i,j})) = \arg.\max(\log(\prod W(E_{i,j}))) = \arg.\max(\sum \log(W(E_{i,j}))) \quad (2)$$

A Eq. 2 mostra que encontrar o caminho que maximiza o produto das probabilidades das arestas é equivalente a encontrar o maior caminho em um grafo onde o peso de cada aresta é igual ao oposto do logaritmo de sua probabilidade, problema conhecido como *Longest Path Problem*, que não apresenta solução em tempo polinomial. No

entanto, como as probabilidades das arestas assumem valores no intervalo  $[0, 1]$ , o logaritmo das arestas (Eq. 2) sempre resulta em números negativos, permitindo a aplicação de um algoritmo clássico de caminhos mínimos em grafos (como Dijkstra), que pode ser resolvido em tempo polinomial. Dessa forma pode ser encontrado o caminho mais provável  $R$  no grafo de grupos.

Nesse ponto vale ressaltar que para encontrar a rota mais provável é necessário que a informação sobre os encontros dos grupos esteja centralizada em algum ponto. Isso não é problema em redes celulares D2D, já que esse processamento pode ser realizado na estação base. Para isso basta que os dispositivos atualizem de forma periódica (e.g. semanalmente) a estação base com os grupos aos quais pertenceram e o número de vezes que esses grupos se encontraram. Dessa forma, no momento em que um usuário desejar enviar um conteúdo para um determinado destinatário, a estação base envia para este usuário a rota grupo-a-grupo mais provável  $R$ . De posse da rota mais provável  $R$ , a cada encontro, um dispositivo que possui a mensagem a encaminha caso o dispositivo encontrado pertença a um dos grupos de  $R$ . Dessa forma a mensagem é propagada grupo-a-grupo, ao longo da rota  $R$ , até o destino.

Embora um caminho grupo-a-grupo seja o mais provável, não há garantia de que esse caminho de fato ocorra. Por isso foi adicionado ao algoritmo o parâmetro  $NRP$ . Com  $NRP = 1$  apenas o caminho mais provável será utilizado para encaminhar as mensagens. Com  $NRP = 2$ , os dois caminhos, independentes em termos de arestas, serão utilizados para o encaminhamento das mensagens. O parâmetro  $NRP$  adiciona um *trade-off* entre taxa de entrega e sobrecarga da rede. Quanto menor  $NRP$  mais seletivo é o encaminhamento. A medida que o parâmetro aumenta, o algoritmo de roteamento tende a se assemelhar a uma inundação (*flooding*).

## 5. Experimentos

### 5.1. Metodologia

Para validar o algoritmo de encaminhamento proposto, realiza-se a comparação com o BubbleRAP [Hui et al. 2011], que se destacou como o mais eficiente em termos de sucesso na entrega das mensagens e baixa sobrecarga da rede. Nos experimentos realizados neste trabalho foram utilizados os parâmetros  $T = 3$  semanas e  $NRP = 1$ . Avaliação do efeito da variação desses parâmetros é considerada nos trabalhos futuros (Seção 6).

#### 5.1.1. Métricas

Com o objetivo de analisar comparativamente o desempenho do algoritmo proposto com relação ao BubbleRAP, que é o atual estado da arte, propõe-se a utilização as seguintes métricas:

- **Taxa de Entrega:** Avalia a porcentagem das mensagens que são entregues com sucesso ao longo do tempo  $t$ .
- **Número de Transmissões:** Mede a sobrecarga da rede, ou seja, o número de transmissões dispositivo-dispositivo que cada algoritmo realiza ao longo do tempo  $t$ .

Nos testes são sorteadas 500 tuplas (*origem, destino, tempo*) e cada um dos algoritmos é executado com cada uma das tuplas sorteadas. Dessa forma, captura-se padrões

de comportamento diversificados ao longo dos traces, conferindo generalidade aos testes. Juntamente com os resultados dos algoritmos são plotados os resultados para uma inundação, em que um nó que possui a mensagem sempre a propaga para um nó que não a possui. A inundação consiste no limite superior para taxa de entrega e sobrecarga da rede. Cada campanha de 500 execuções é repetida oito vezes com diferentes sementes para geração das 500 tuplas aleatórias. Com isso, foram obtidos intervalos confiança de 95% para os testes.

### 5.1.2. BubbleRAP

O algoritmo BubbleRAP identifica comunidades sociais densamente interconectadas a partir do grafo de contatos agregados de todo o trace. Portanto, cada nó passa a pertencer a pelo menos uma comunidade. Atribui-se a nós que não estejam em nenhuma comunidade coletiva uma pseudo-comunidade de apenas um nó, para fins de encaminhamento. Além disso, a cada nó é atribuída uma medida de sua popularidade em relação a todos os nós do trace (*GlobalRank*) e em relação somente aos nós membros de sua comunidade (*LocalRank*).

Com isso o encaminhamento da origem até o destinatário final se dá de acordo com a seguinte política:

- A cada encontro, um dado nó transmite a mensagem se o nó encontrado tiver um *GlobalRank* maior que o seu próprio ou se o nó encontrado pertencer a uma comunidade em conjunto com o destinatário final.
- Uma vez que a mensagem chega a algum nó membro da comunidade à qual o destinatário final pertence, o encaminhamento fora desta comunidade deixa de ocorrer.
- Uma vez que a mensagem chegou na comunidade destino, o encaminhamento passa a ocorrer caso o *LocalRank* do nó encontrado seja maior que o do nó que possui a mensagem ou caso o nó encontrado seja o destinatário final da mensagem.

Para fins de comparação, o BubbleRAP foi implementado utilizando os parâmetros reportados em [Hui et al. 2011] e que geraram os melhores resultados em termos de taxa de entrega e sobrecarga da rede. O cálculo do *GlobalRank* e do *LocalRank* foi realizado utilizando a técnica C-Window que, segundo [Hui et al. 2011], melhor aproximou métrica de centralidade de um nó em seus experimentos.

## 5.2. Resultados

As Figuras 5 e 6 apresentam, em escala logarítmica, os resultados em termos de taxa de entrega e sobrecarga da rede com transmissões em função do tempo após o início da propagação da mensagem.

Os resultados para o *trace* MIT Reality Mining, na Figura 5, mostram que nas primeiras horas após o início da propagação o GROUPS-NET apresenta uma taxa de entrega sutilmente maior e nas horas finais Bubble Rap ultrapassa o GROUPS-NET. Além disso, ao longo de todo o tempo de propagação, o GROUPS-NET apresentou uma sobrecarga pouco maior que a do Bubble Rap. Vale ressaltar que o *trace* MIT Reality Mining é muito particular, pois todos os nós monitorados residem no mesmo local. Logo, espera-se que os laços sociais desse subconjunto dos usuários sejam mais fortes do que a média, fato esse que beneficia o Bubble Rap, que usa a estrutura social estática e a popularidade dos nós

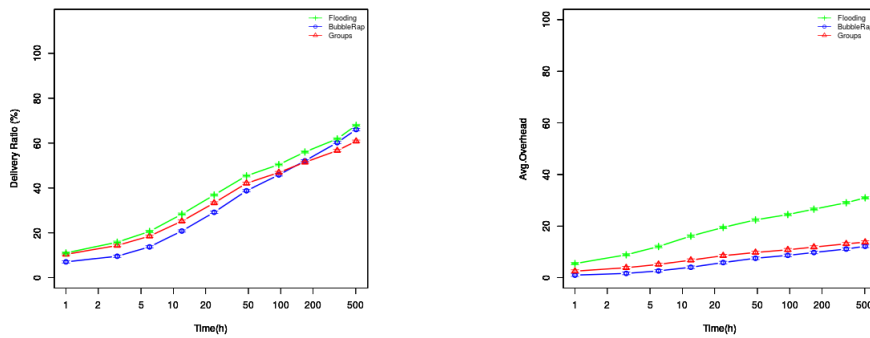


Figura 5. MIT Reality (80 nós): Taxa de entrega e número de transmissões.

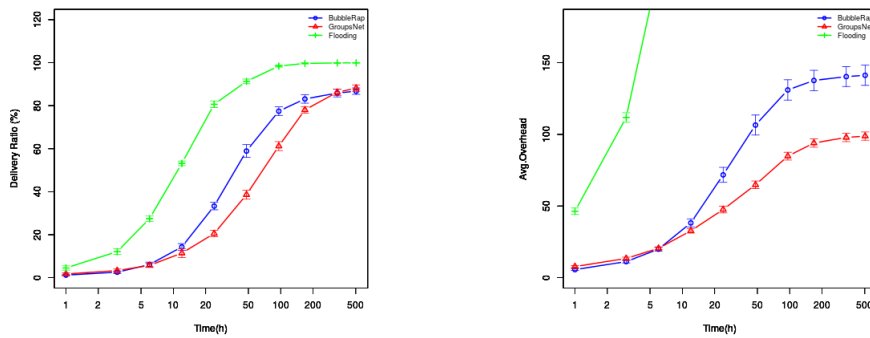


Figura 6. Dartmouth (1200 nós): Taxa de entrega e número de transmissões.

para definir sua política de encaminhamento. Esse fato que motivou a realização de testes em um trace de larga escala como o Dartmouth. Não obstante, é importante enfatizar que, mesmo nesse cenário específico, o GROUPS-NET apresentou um desempenho bastante similar ao do Bubble Rap, com a vantagem de não requerer detecção de comunidades.

A Figura 6 mostra os resultados obtidos para o trace Dartmouth. No *trace* Dartmouth, 1200 alunos de um campus universitário foram monitorados, sendo portanto um trace de mais larga escala e menos específico que o MIT Reality Mining. Nesse caso o GROUPS-NET obteve um desempenho consideravelmente melhor que o Bubble RAP. No período entre 24 e 96 horas após o início da transmissão o Bubble Rap obtém uma maior taxa de entrega, sendo ultrapassado pelo GROUPS-NET nas horas seguintes e até o fim do período de três semanas. Em relação à sobrecarga da rede, após a sexta hora o Bubble Rap passa a transmitir muito mais mensagens que o GROUPS-NET, apresentando uma sobrecarga em média 50% maior que a do GROUPS-NET nas horas seguintes. Para mil mensagens (*origem, destino*) no trace Dartmouth, isso representa uma economia de cerca de 60 mil transmissões dispositivo-dispositivo.

Definindo-se uma métrica de benefício-custo como a razão entre a taxa de entrega e a sobrecarga da rede a cada instante de tempo, o GROUPS-NET mantém um desempenho equivalente ou melhor que o Bubble RAP em todos os momentos após o início das transmissões. Como mostrado na figura 7, após as primeiras horas de transmissão, o desempenho do GROUPS-NET, em termos de benefício-custo, chega a ser 80% maior. Como mencionado anteriormente, o GROUPS-NET ainda tem a vantagem de utilizar detecção de grupos e não detecção de comunidades, o que é mais viável na prática.

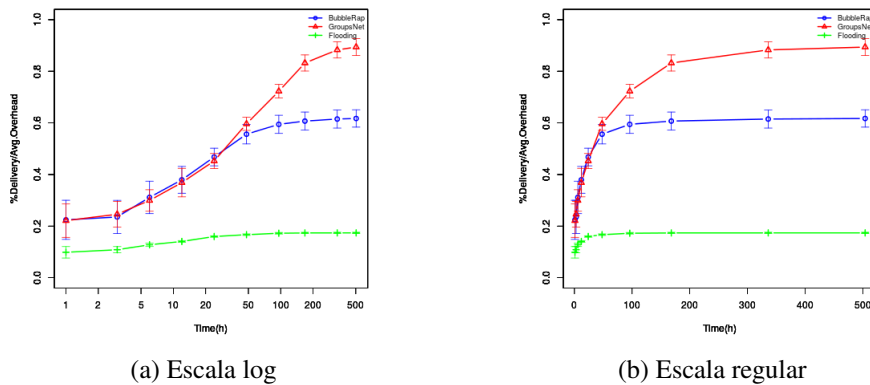


Figura 7. Comparação da relação benefício-custo dos protocolos Bubble Rap e Groups ao longo do tempo no trace Dartmouth.

## 6. Conclusões e Trabalhos Futuros

Neste trabalho foi descrito o GROUPS-NET, um algoritmo de roteamento oportunístico ciente de encontros de grupos para redes celulares D2D. O algoritmo possui a vantagem de não necessitar de detecção distribuída de comunidades, além de considerar a dinamicidade de encontros sociais em contraste com estratégias anteriores, que consideram a estrutura social estática de comunidades.

O GROUPS-NET foi comparado com o algoritmo do estado da arte Bubble Rap em um cenário particular e de pequena escala, por meio do trace MIT Reality, apresentando desempenho similar. Já em um cenário de larga escala e menos específico, como o *trace* de Dartmouth, o GROUPS-NET demonstrou um desempenho superior, realizando 40% menos transmissões para atingir a mesma taxa de entrega. Foi analisada a relação benefício-custo dos dois algoritmos e para todos os tempos de transmissão o GROUPS-NET apresentou melhor benefício-custo.

Como trabalho futuro destaca-se a utilização de informações de geo-localização e pontos de interesse combinados com a ciência de encontros de grupos para aprimorar os resultados do algoritmo. Além disso, pode ser feito um estudo de como a variação no período de tempo a ser considerado antes da transmissão e o número de caminhos redundantes (parâmetros  $T$  e  $NRP$ ) afetam o desempenho do algoritmo em cada cenário de testes. Finalmente, destaca-se a possibilidade de adaptar o GROUPS-NET para atuar em cenários que a mesma mensagem tem múltiplos destinatários.

## Referências

- Bauza, R., Gozalvez, J., and Sanchez-Soriano, J. (2010). Road traffic congestion detection through cooperative vehicle-to-vehicle communications. In *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, pages 606–612. IEEE.
- Calabrese, F., Pereira, F. C., Di Lorenzo, G., Liu, L., and Ratti, C. (2010). The geography of taste: analyzing cell-phone mobility and social events. In *Pervasive computing*, pages 22–37. Springer.
- Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268.

- Fodor, G., Dahlman, E., Mildh, G., Parkvall, S., Reider, N., Miklós, G., and Turányi, Z. (2012). Design aspects of network assisted device-to-device communications. *Communications Magazine, IEEE*, 50(3):170–177.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.
- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10):103018.
- Henderson, T., Kotz, D., and Abyzov, I. (2008). The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712.
- Hui, P., Crowcroft, J., and Yoneki, E. (2011). Bubble rap: Social-based forwarding in delay-tolerant networks. *Mobile Computing, IEEE Transactions on*, 10(11):1576–1589.
- Hui, P., Yoneki, E., Chan, S. Y., and Crowcroft, J. (2007). Distributed community detection in delay tolerant networks. In *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*, page 7. ACM.
- Lei, L., Zhong, Z., Lin, C., and Shen, X. (2012). Operator controlled device-to-device communications in lte-advanced networks. *IEEE Wireless Communications*, 19(3):96.
- Li, Y., Wu, T., Hui, P., Jin, D., and Chen, S. (2014). Social-aware d2d communications: qualitative insights and quantitative analysis. *Communications Magazine, IEEE*, 52(6):150–158.
- Lindgren, A., Doria, A., and Schelén, O. (2003). Probabilistic routing in intermittently connected networks. *ACM SIGMOBILE mobile computing and communications review*, 7(3):19–20.
- Lu, J. and Cao, L. (2003). Congestion evaluation from traffic flow information based on fuzzy logic. In *Intelligent Transportation Systems, 2003.*, volume 1, pages 50–53. IEEE.
- Nguyen, N. P., Dinh, T. N., Tokala, S., and Thai, M. T. (2011). Overlapping communities in dynamic networks: their detection and mobile applications. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 85–96. ACM.
- Nunes, I., Vaz de Melo, P., and A.F. Loureiro, A. (2016). Group mobility: Detection, tracking and characterization. In *IEEE ICC 2016 International Conference on Communications (ICC'16 SAC-8 SN)*, Kuala Lumpur, Malaysia.
- Palla, G., Barabási, A.-L., and Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136):664–667.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.
- Peel, L. (2010). Estimating network parameters for selecting community detection algorithms. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE.
- Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., and Loureiro, A. A. (2014). Revealing the city that we cannot see. *ACM Transactions on Internet Technology (TOIT)*, 14(4):26.
- Xu, H., Hu, Y., Wang, Z., Ma, J., and Xiao, W. (2013). Core-based dynamic community detection in mobile social networks. *Entropy*, 15(12):5419–5438.