

MOCHA: Um framework para caracterização e comparação de traces de mobilidade

Fabício R. Souza¹, Augusto C. S. A. Domingues¹,

Pedro O. Vaz de Melo¹, Antonio A. F. Loureiro¹

¹Departamento de Ciência da Computação - Universidade Federal de Minas Gerais

{fabriciors, augusto.souza, olmo, loureiro}@dcc.ufmg.br

Abstract. *There are many mobility models in the literature, with diverse formats as well as its origins. Besides the existence of works that analyse and characterize these models, there is a need for a framework that can compare them in an easy manner. MOCHA (Mobility framework for CHaracteristics Analysis) is a tool that characterizes and makes possible the comparison of mobility models without any hard work. We implemented 9 social, spatial and temporal characteristics, which were extracted from various (real and synthetic) different mobility traces. The metrics used in the tool can become a standard for trace analysis and comparison in the literature, allowing a better vision of where one trace belongs related to others.*

Resumo. *Existem diversos modelos de mobilidade na literatura, em formatos diversos bem como suas origens. Apesar da existência de diversos trabalhos que caracterizem esses modelos, há a necessidade de um framework que possa compará-los de maneira fácil. O MOCHA é uma ferramenta que caracteriza e possibilita a comparação de modelos de mobilidade de forma simples. Foram implementadas 9 características sociais, espaciais e temporais, e essas características foram extraídas de diversos traces (reais e sintéticos) de mobilidade diferentes. A ferramenta provê diversas métricas que podem se tornar padrão para a comparação de traces e permite um entendimento de onde um trace se encontra em relação aos outros.*

1. Introdução

Existem diversos modelos de mobilidade na literatura com formatos, origens e propósitos diversos. Tais modelos possuem aplicações importantes na simulação de redes móveis, representando como seus usuários se comportam, e portanto a sua proximidade da realidade é um fator indicativo da qualidade do modelo. *Small World in Motion* (SWIM) [Mei e Stefa 2009], *Self-similar Least Action Walk* (SLAW) [Lee et al. 2009] e *Working Day Model* (WDM) [Ekman et al. 2008a] são modelos frequentemente aplicados na literatura. SWIM é um modelo que se baseia em métricas de redes complexas para criação de grafos "small-world" entre os usuários. SLAW foi um dos primeiros modelos de mobilidade a considerar distribuições estatísticas como leis de potência truncada para as viagens de usuários, tempos de espera, etc. Por fim, WDM é um modelo que se baseia no dia de trabalho tradicional de uma pessoa: ela acorda no mesmo horário, vai trabalhar no mesmo local e ao fim do dia, retorna para casa. Apesar de todos eles representarem a mobilidade

de usuários em uma rede, diferentes contextos e modelagens são considerados em cada um, o que dificulta uma comparação direta entre eles.

A capacidade de comparar modelos é importante. Por exemplo, podemos comparar um *trace* sintético gerado usando um determinado modelo de mobilidade com um *trace* real para verificar o quão próximo da realidade as amostras geradas estão, o que é um fator direto de qualidade. No sentido contrário, comparar um *trace* real com *traces* sintéticos gerados a partir de modelos de mobilidade conhecidos, nos permite identificar com qual modelo de mobilidade este *trace* se parece. Essa informação fornece uma visão geral dos dados, permitindo análises e tomadas de decisão sem que exista a necessidade do processamento dos dados reais.

Um dos fatores que dificulta a comparação entre *traces* de diferentes tipos são os dados presentes em cada um. Em *traces* de contato (como DARTMOUTH [Kim et al. 2006] e SWIM [Mei e Stefa 2009]), as amostras representam encontros entre pares de usuários na rede em uma determinada localidade, que pode ser uma coordenada geográfica ou um ponto de interesse. Esse tipo de *trace* é conhecido por ser temporalmente esparsos, isto é, não se sabe exatamente as ações tomadas por um usuário entre um contato e outro. Já em *traces* GPS (como [Piorkowski et al. 2009]), as amostras representam a posição (coordenadas de longitude e latitude) precisa de um usuário em um determinado instante de tempo. Porém, não é claro a ocorrência de contatos entre os usuários: essa informação precisa ser minerada do *trace*. Por fim, *traces* de check-in (como [Yang et al. 2015]) representam a confirmação de presença de usuários em um determinado ponto de interesse. Assim, vemos que a precisão espacial, a granularidade temporal e informações contextuais dificultam a comparação direta entre esses tipos.

Com isso, há a necessidade de uma maneira de realizar essas comparações. O MOCHA (Mobility framework for CHaracteristics Analysis) é uma ferramenta que permite a extração de características de *traces* de mobilidade em diferentes formatos, facilitando assim uma comparação entre eles. A ferramenta é dividida em quatro módulos: o primeiro, converte os *traces* de diferentes formatos para um formato padrão definido, sem que haja perda de informação durante a conversão; o segundo extrai 9 características temporais, espaciais e sociais; o terceiro classifica cada uma das características de acordo com sua distribuição estatística; e o quarto módulo utiliza o método T-SNE (*t-Distributed Stochastic Neighbor Embedding*) para comparar e agrupar os *traces* de acordo com os distribuições obtidas das características.

Como validação da ferramenta, *traces* reais e sintéticos de diferentes tipos de mobilidade existentes na literatura foram comparados. Os resultados obtidos são apresentados e neles é possível notar a aproximação dos *traces* com características semelhantes.

O restante deste trabalho está organizado da seguinte maneira. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3, as características extraídas pelo MOCHA são definidas. A Seção 4 apresenta a ferramenta. Por fim, as Seções 5 e 6 apresentam os resultados da análise de *traces* da literatura e as conclusões obtidas.

2. Trabalhos Relacionados

A caracterização de modelos e *traces* de mobilidade está presente em muitos trabalhos na literatura. Em [Thakur e Helmy 2013], os autores apresentam um método de *benchmarking* de mobilidade, incluindo um espaço métrico multi-dimensional de mobilidade

e um *framework* para a análise de mobilidade e desempenho de protocolos em *traces*. Adicionalmente, um novo modelo de mobilidade é proposto e analisado juntamente com outros modelos da literatura, usando o *framework* proposto. COBRA (*COllective Behavior based on Realistic Aspects of human mobility*), é um modelo de mobilidade que visa capturar métricas individuais, par-a-par, e em grupo. Apesar de apresentar um *framework* para a análise de diferentes modelos e *traces* de mobilidade, o trabalho não considera a classificação de *traces* em relação às suas características.

Os autores de [Bai et al. 2003] propõem IMPORTANT (Impact of Mobility on the Performance Of Routing protocols in Adhoc NeTworks), um *framework* para analisar o impacto da mobilidade no desempenho de protocolos de roteamento em redes *ad hoc*, focando em modelos de mobilidade, métricas de mobilidade, e propriedades de grafos em redes. Entretanto, são poucos os modelos de mobilidade considerados.

Em [Munjal et al. 2010] são apresentados modelos para facilitar a construção de cenários de simulação rigorosa de MANETs. Para isso, os autores consideram três métricas para a avaliação do desempenho de protocolos de roteamento em MANETs: *average shortest-path hop count*, *average network partitioning*, e *average neighbor count*. Apesar de apresentar métricas genéricas para a descrição da mobilidade em redes móveis, fatores sociais e temporais, que são de extrema relevância em redes móveis, não são considerados nesse trabalho.

O SLkit é apresentado em [Ferreira et al. 2017], uma ferramenta capaz de automatizar o processo de extração e análise de métricas espaciais e temporais em *Sensing Layers*, com o intuito de agilizar trabalhos que consistem da investigação de múltiplos *datasets*. Contudo, a ferramenta não considera a extração de características sociais dos *traces*.

[Bezerra et al. 2009] investigam o comportamento dos usuários e sua mobilidade em relação à mesma modelada através de modelos sintéticos de mobilidade, com o intuito de validar tais modelos quanto ao nível de proximidade da realidade. Além disso, propõem uma técnica de ajuste dos parâmetros dos modelos de mobilidade baseada nos valores obtidos de amostras reais. Similarmente, [Baumann et al. 2008] propõem um *framework* para a simulação de mobilidade genérica, que permite comparar modelos de mobilidade considerando métricas espaciais, temporais e sociais. O trabalho considera os modelos *Random Waypoint Model*, *Manhattan Model*, e modelos *GIS*.

A partir da análise da literatura, é possível observar que existe a necessidade da implementação de um *framework* que permita a comparação de diferentes *traces* e modelos de mobilidade. O MOCHA é capaz de comparar *traces* de diferentes formatos, como GPS, contatos e check-in, extraíndo métricas de mobilidade que consideram características temporais, espaciais e sociais. A comparação visual oferecida considera as distribuições estatísticas e seus parâmetros estimados para cada métrica. Assim, é possível agrupar *traces* semelhantes, o que oferece *insights* sobre um espaço cartesiano de modelos de mobilidade. Isto permite a avaliação de qualidade de *traces* sintéticos gerados. A disponibilidade da implementação desse *framework* à comunidade científica facilita o trabalho necessário para a análise de modelos de mobilidade, como também contribui para uma padronização das métricas utilizadas para a caracterização de *traces* reais e sintéticos.

3. Métricas

Traces de mobilidade podem ser usados para inúmeras aplicações, que podem ser mais sensíveis a diferentes tipos de métricas. Um dos tipos de métricas são as espaciais, que descrevem como um usuário se relaciona com o espaço que ele ocupa. Outro tipo de métrica são as temporais, que apresentam informações sobre a relação do usuário com o tempo gasto em tarefas do seu dia-a-dia. Existem também as métricas sociais, que ilustram a relação de um usuário com os outros usuários da rede. Considerando as diferenças levantadas acima, é interessante considerar características de diferentes contextos para a análise e comparação de modelos de mobilidade.

Assim, esta seção apresenta as métricas aplicadas na caracterização dos *traces*.

3.1. Espaciais

3.1.1. Raio de Giro (*RDG*)

Na mobilidade social, é comum supor que cada nó possui uma localidade considerada como sua casa. [Kosta et al. 2014] definem a casa como um ponto aleatório escolhido uniformemente na região da rede. [Ekman et al. 2008b] a consideram como o ponto de início das atividades diárias de cada nó. Para o propósito deste trabalho, consideramos a casa como sendo o ponto mais visitado na rotina de cada nó, onde normalmente o usuário retorna no fim do dia.

Podemos assumir que um nó tende a se mover para localidades próximas a sua casa para realizar compromissos, fazer compras, visitar amigos, etc. Existe uma propriedade da mobilidade que mede a distância que um nó orbita ao redor de sua casa. Chamamos essa distância de Raio de Giro. Assim, o Raio de Giro de um usuário é a distância euclidiana máxima percorrida por ele, partindo de sua casa para uma localidade.

3.1.2. Distância de Viagem (*DV*)

Quando um nó se move de uma localidade para outra, descrevemos esta ação como uma viagem, salto ou vôo [Karamshuk et al. 2011, Ekman et al. 2008b]. A Distância de Viagem é definida como sendo a distância percorrida entre duas localidades consecutivas [Ekman et al. 2008b].

A *DV* relaciona-se com o *RDG* sendo que a principal diferença entre elas é o fato de que o *RDG* considera as distâncias de viagem de acordo com a casa do usuário, e não a distância entre localidades visitadas consecutivamente. A análise dessas propriedades permite entender como os nós se movem de um lugar para outro. Essa informação pode ser explorada, por exemplo, para melhorar a disseminação de mensagens em redes veiculares e oportunísticas.

3.2. Temporais

3.2.1. Tempo de Visita (TV_s)

O Tempo de Visita pode ser definido como o tempo gasto em cada localidade visitada por um nó [Karamshuk et al. 2011].

3.2.2. Tempo de Viagem (TV_g)

De forma análoga à Distância de Viagem, o Tempo de Viagem é definido como o tempo gasto por um nó para se mover de uma localidade à outra.

3.3. Sociais

3.3.1. Tempo de Inter-Contato (TIC)

O Tempo de Inter-Contato pode ser definido como o intervalo de tempo entre encontros consecutivos de um par de nós dentro do trace [Kosta et al. 2014]. Apesar do TIC representar o tempo entre contatos consecutivos de um mesmo par de nós, alguns autores, como [Helgason et al. 2014], preferem usar uma definição mais genérica chamada Tempo de Inter-Contato-Qualquer, que considera contatos consecutivos entre qualquer par de agentes. Neste trabalho foi utilizada a primeira definição, por ser mais popular na literatura e representar melhor a periodicidade de encontros na realidade. Em cenários oportunistas o TIC fornece informações sobre a frequência com que cada par de agentes se encontram, representando uma oportunidade de transmissão de mensagem [Song e Kotz 2007].

3.3.2. Duração de Contato (DC)

A definição da duração de um contato é simples ela é a quantidade de tempo que um par de agentes está localizado dentro do alcance de comunicação um do outro, sem interrupções. De forma análoga ao TIC , DC também representa uma oportunidade para transmitir uma mensagem em uma rede oportunística. Entretanto, enquanto o TIC fornece informações sobre quão frequente uma mensagem pode ser enviada, DC permite compreender por quanto tempo um contato ocorre, e portanto, a quantidade de dados que pode ser transferida durante o mesmo. O conhecimento dessa propriedade está diretamente relacionado à possível vazão que cada nó pode transmitir enquanto em contato com outros nós.

3.3.3. Regularidade de Encontro (RE)

Regularidade de Encontro é uma métrica de redes complexas que mapeia a regularidade de uma relação social. É possível afirmar que um nó tende a se encontrar mais frequentemente com seus amigos ou conhecidos, enquanto contatos aleatórios raramente acontecem [de Melo et al. 2015]. Quanto mais contatos entre cada par de nós em uma janela de tempo específico, mais provável é a existência de um laço social entre esses nós. Ao considerar cenários oportunistas, a Regularidade de Encontro representa o quão frequente será possível entregar uma mensagem entre um par específico de nós.

3.3.4. Sobreposição Topológica (ST)

Outra propriedade que fornece informações sobre a rede social em cada cenário é a Sobreposição Topológica. ST representa a sobreposição social existente entre cada par de nós quando consideramos todos os encontros nos quais eles participaram. ST pode ser

usada para determinar oportunidades para entregar mensagens dentro de uma comunidade de nós. Intuitivamente, é mais provável entregar uma mensagem para um membro de uma mesma comunidade social do que para um completo estranho.

3.3.5. Correlação Social (CS)

A Correlação Social é o valor da correlação de Pearson entre os valores de Sobreposição Topológica e Regularidade de encontros. Este valor representa o quão estável é a relação entre as pessoas que um usuário encontra enquanto se move e a quantidade de vezes que ele viu essas pessoas num determinado intervalo de tempo. Essa métrica pode ser utilizada para verificar se seria eficiente a disseminação de dados para um grupo de pessoas, se possuem CS alta significa que o mesmo grupo se encontra frequentemente.

4. MOCHA

Neste trabalho é apresentado o MOCHA, uma ferramenta para processamento, extração, classificação e comparação de características. A Figura 1 ilustra as etapas executadas pelo MOCHA, com as entradas necessárias e as saídas produzidas por cada uma. As subseções seguintes descrevem em detalhes cada um desses módulos.

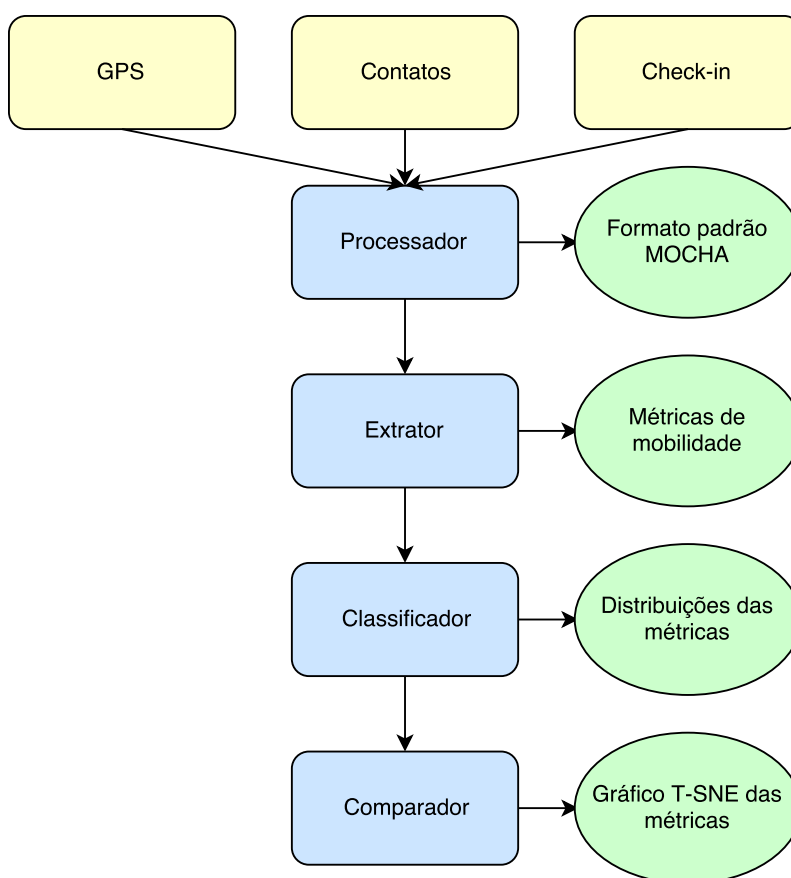


Figura 1. Processo de funcionamento do MOCHA

4.1. Processamento

Existem diversos *traces* de mobilidade na literatura, como SWIM [Mei e Stefa 2009], SLAW [Lee et al. 2009], SMOOTH, DARTMOUTH [Kim et al. 2006], entre outros. Devido ao contexto da coleta, por exemplo, quais dispositivos foram usados para a coleta e qual o intuito da mesma, cada um destes *traces* possui um formato diferente para registrar a mobilidade de seus agentes, o que se torna um problema para a extração das métricas.

O módulo de processamento do MOCHA é responsável pela normalização destes *traces* num formato único. O formato utilizado é ID1 ID2 TA TI TD X1 Y1 X2 Y2 onde:

- *ID1* é um número inteiro único para o Usuário 1;
- *ID2* é um número inteiro único para o Usuário 2;
- *TA* é o tempo atual do *trace*;
- *TI* é o tempo em que o contato entre *ID1* e *ID2* se iniciou;
- *TD* é o tempo de duração do contato entre *ID1* e *ID2*;
- *X1* é a abscissa do Usuário 1;
- *Y1* é a ordenada do Usuário 1;
- *X2* é a abscissa do Usuário 2;
- *Y2* é a ordenada do Usuário 2;

Neste formato estão representados todos os encontros que aconteceram durante a coleta do *trace*.

Para definir os encontros, o MOCHA possui um parâmetro que é o raio de comunicação, que é utilizado para identificar quando dois usuários estão próximos o suficiente para poderem trocar informações. Assim, toda vez em que um usuário se movimenta, são verificados quais usuários ainda estão no raio de comunicação deste, quais usuários deixaram de estar no raio de comunicação e quais usuários entraram nesse raio. Quando dois usuários que estavam em comunicação deixam de se comunicar, é gerada uma nova entrada no formato MOCHA.

4.2. Extração

Uma vez que os dados estejam no formato MOCHA, podemos proceder para a extração das propriedades desejadas. O objetivo do extrator é extrair do *trace* todas as propriedades de mobilidade descritas na Seção 3. Após a extração, podemos analisar a curva estatística gerada pela agregação de todas as ocorrências de uma propriedade, por exemplo.

É importante colocar em evidência algumas suposições que devem ser consideradas. Primeiro, nem todos os *traces* analisados pelo MOCHA possuem todas as características possíveis para extração. Em um *trace* de check-in sem as coordenadas geográficas dos pontos de interesse, é impossível calcular propriedades como Distância de Viagem. Similarmente, um *trace* de contatos sem coordenadas geográficas permite somente a extração das métricas sociais, excluindo aquelas temporais e espaciais. Entretanto, é sempre possível extrair todas as métricas sociais.

Adicionalmente, existem algumas propriedades nas quais sua extração depende de parâmetros de configuração sensíveis ao contexto da simulação. Por exemplo, todos os contatos são avaliados com base em um raio de 50 metros. Porém, esse parâmetro pode ser ajustado de acordo com o contexto, se necessário. O valor de 50 metros é considerado o padrão neste trabalho pois *traces* amplamente usados na literatura, como

Tabela 1. Distribuições estatísticas atribuídas às métricas extraídas dos *traces*

Trace	Métricas								
	RDG	DV	TIC	DC	RE	ST	CS	TV_s	TV_g
RAW	Normal	Normal	Normal	Normal	Exponencial	Exponencial	0.016	Pareto	Exponencial
SLAW	Exponencial	Normal	Normal	Normal	Exponencial	Normal	0.020	Pareto	Exponencial
SWIM	Normal	Normal	Normal	Normal	Exponencial	Exponencial	0.017	Pareto	Exponencial
WDM	Exponencial	Normal	Normal	Normal	Exponencial	Exponencial	0.552	Log-normal	Pareto
SAN FRANCISCO	Normal	Normal	Normal	Normal	Exponencial	Pareto	0.43	Pareto	Pareto
DARTMOUTH	Exponencial	Normal	Normal	Normal	Exponencial	Normal	0.38	Pareto	Exponencial

Dartmouth [Kim et al. 2006], são baseados na proximidade de roteadores *Wi-fi*, onde um contato é considerado sempre que dois usuários distintos estão conectados a um mesmo roteador. É comum na literatura definir o raio de alcance de redes *Wi-fi* como sendo aproximadamente 50 metros [Dimatteo et al. 2011].

Finalmente, considerando que o processo de extração é executado somente para *traces* processados no formato MOCHA, todos os *traces* devem estar ordenados cronologicamente e cada entrada deve estar no formato descrito na Seção 4.1.

4.3. Classificação e comparação

Com as características extraídas é possível classificá-las de acordo com suas distribuições estatísticas. Para classificar as características foram utilizados os métodos de *Maximum Likelihood Estimation* (MLE) [Scholz 2004] em conjunto com o *Akaike Information Criteria* (AIC) [Akaike 2011].

O MLE é um método que identifica o quão próximo de uma dada distribuição um conjunto de dados está. O erro quadrático entre os dados e a distribuição são calculados e a distribuição que possui o menor erro quadrático é a que mais se aproxima dos dados originais. No entanto, distribuições diferentes utilizam um conjunto de parâmetros diferentes, o que é de extrema importância para a classificação. Não é possível considerar que uma distribuição que utilize vários parâmetros a mais que outra possua o mesmo nível de detalhes na classificação. Para resolver este problema, utilizamos o AIC, que utiliza a Equação 1 para decidir qual a melhor distribuição.

Na Equação 1, E representa o erro quadrático calculado pelo MLE e P representa a quantidade de parâmetros que a distribuição utiliza. A distribuição que possui o menor valor de AIC é a distribuição que mais se aproxima dos dados reais. O MOCHA compara todas as características extraídas, exceto a CS , com sete distribuições diferentes, que são: Gamma, Weibull, Exponencial, Normal, Lognormal, Pareto (power-law) e Loglogistic [John Lu 2010]. As classes de distribuições foram escolhidas de forma a cobrir a maior variação possível com a menor quantidade de testes, com o intuito de diminuir o tempo de processamento. A Tabela 1 apresenta um exemplo de como as características extraídas são classificadas. É possível ver na métrica ST que *traces* diferentes podem apresentar classificações do mesmo parâmetro diferentes, e o mesmo ocorre para a métrica CS, onde o menor valor é próximo de zero enquanto o maior valor passa de 0,5, demonstrando grandes diferenças nos *traces* analisados.

$$-\text{Log}(E) + 2(P + 1) \quad (1)$$

4.3.1. Comparação

Feita a classificação das métricas de acordo com suas distribuições estatísticas, proseguimos para a etapa de comparação dos *traces*. Para isso, usamos a classificação das métricas com o intuito de verificar quais *traces* possuem semelhanças ou diferenças entre si. Esta informação é de suma importância para os estudos que usam *traces* de mobilidade como tema. Primeiro, é possível verificar o quão semelhante um novo *trace* (real ou sintético) é em relação àqueles usados na literatura. Isto permite aos pesquisadores assumir conhecimentos obtidos anteriormente a partir de *traces* conhecidos a novos *traces*, reduzindo assim o tempo necessário com análises e simulações. Segundo, também é possível avaliar a qualidade de novos *traces* (reais ou sintéticos) através da comparação. Por exemplo, ao gerar um conjunto de dados baseado em um certo modelo de mobilidade da literatura, espera-se que este conjunto de dados seja semelhante àqueles que também foram derivados desse mesmo modelo. Por fim, o uso dessa ferramenta, em conjunto com a disponibilização dos dados por parte da comunidade, permite a elaboração de um "mapa" de caracterização de *traces*, que pode ajudar nas escolhas dos conjuntos de dados a serem usados em estudos.

Para realizar a comparação, utilizamos a técnica t-SNE (*t-Distributed Stochastic Neighbor Embedding*) [Maaten e Hinton 2008], que possibilita a visualização de dados com alta dimensionalidade em um gráfico de duas ou três dimensões mantendo grande parte da informação original, isto é, preservando o agrupamento que existia em uma alta dimensão. O t-SNE produz visualizações dos dados que são significativamente melhores que aquelas produzidas por outras técnicas de agrupamento de dados [Maaten e Hinton 2008], como a Análise de Componentes Principais (*Principal Component Analysis*, ou PCA). Essa foi a razão pela escolha desta técnica, pois a comparação fornecida pelo MOCHA visa, acima de tudo, oferecer um guia visual das relações entre os *traces*.

O Algoritmo funciona da seguinte maneira. Primeiro, ele determina as similaridades entre todos os *traces* analisados, considerando as características sociais, espaciais e temporais como dimensões e suas distribuições de probabilidade como valor. A partir disso, o algoritmo produz um gráfico em baixa dimensão (neste caso, optamos por duas dimensões pela facilidade de entendimento) com os *traces* colocados em posições aleatórias. Então, a cada passo, ele aproxima aos poucos os *traces* similares e distancia os não similares de acordo com suas distâncias estatísticas, atingindo ao fim de sua execução, agrupamentos em baixa dimensão similares àqueles existentes em alta dimensão.

5. Análise

O MOCHA foi utilizado para extrair, classificar e comparar seis *traces* de mobilidade diferentes. Os *traces* foram obtidos de bases de dados próprias de cada projeto e estão disponíveis online. A Figura 2 apresenta os resultados da comparação. Nela, a distância entre os pontos representa a diferença entre eles, ou seja, o quão mais próximos, mais similares dois *traces* são. Podemos observar que os *traces* SWIM e SLAW, que são *traces* sintéticos que utilizam métricas de redes complexas, estão próximos um do outro, enquanto os *traces* Dartmouth, RAW e San Francisco, que são *traces* reais, ficaram próximos entre si. O *trace* WDM, apesar de ser um *trace* sintético, ficou distante dos outros dois sintéticos, e próximo dos reais. Isso demonstra uma habilidade melhor para gerar mobilidade que se assemelhe a uma mobilidade real.

A distancia entre o *trace* Dartmouth e o *trace* San Francisco representa a diferença de comportamento dos usuários. No Dartmouth os agentes são pessoas se movimentando em um campus universitário, ou seja, possuem locais de interesse (como cafeterias, salas de aula, moradias) que se encontram próximos uns aos outros enquanto no San Francisco, que é um *trace* de movimento de motoristas de táxi da cidade, as viagens dos taxistas são maiores do que o movimento que as pessoas do campus executam. Além disso, um motorista de táxi deseja sempre estar em movimento, para que não perca dinheiro, enquanto os alunos possuem tempos de espera maiores, porque estão assistindo aulas, trabalhando em laboratórios, bibliotecas, etc.

O *trace* RAW representa pessoas se movimentando no dia-a-dia, ou seja, trabalhando, executando tarefas, se entretendo, ficando em casa, etc. Podemos ver que ele se encontra num meio termo entre o Dartmouth e o San Francisco. Uma pessoa que não trabalha como taxista não tem viagens tão longas e frequentes quanto os agentes do *trace* de San Francisco, e também não tem viagens frequentes e curtas como os agentes do Dartmouth, porque não passam períodos similares (tempo de uma aula) num local e, em seguida se movimentam para outro. O que demonstra que, por colocar o *trace* como um meio termo entre Dartmouth e San Francisco, o MOCHA conseguiu caracterizar bem a relação entre os comportamentos dos agentes de cada *trace*.

O *trace* SWIM tem como intuito a criação de uma movimentação que gere "*small-worlds*" que se movimentem pelo espaço, ou seja, grupos de agentes que tem uma correlação social alta se movimentam em conjunto, enquanto o WDM se preocupa apenas com o movimento individual dos agentes, considerando o movimento em grupo apenas em casos de eventos sociais. É possível notar essa diferença na distância entre os dois pontos no gráfico.

6. Conclusão

Com a grande quantidade de estudos que utilizam *traces* como fontes de dados para simulações, fez-se necessário avaliar a qualidade desses conjuntos de dados presentes na literatura. Duas maneiras convencionais de se avaliar a qualidade de um *trace* são a extração de métricas pré-definidas e a comparação com outros *traces* já conhecidos. Porém, não existe um padrão de métricas a serem extraídas para comparação. Adicionalmente, os diferentes formatos de *traces* existentes na literatura (contatos, GPS, check-in, Wi-fi, Bluetooth) dificultam uma comparação direta entre os mesmos.

Neste trabalho, apresentamos o MOCHA, um *framework* capaz de extrair métricas de *traces* de mobilidade. Primeiro, ele é capaz de converter *traces* em diferentes formatos (contatos, GPS, check-in) para um formato padrão definido. A partir destes dados convertidos, métricas sociais, espaciais e temporais são extraídas e classificadas quanto à sua distribuição de probabilidade. Por fim, o MOCHA compara os *traces* estudados de acordo com as distribuições de suas métricas, fornecendo um guia visual de como eles se assemelham. Esta comparação pode ser usada para avaliar a qualidade de um novo *trace*, real ou sintético. A semelhança entre *traces* pode reduzir o número de simulações em um estudo, por exemplo.

Trabalhos futuros incluem permitir a geração de dados sintéticos através da definição manual das métricas, de modo a gerar um *trace* que se encaixe nas necessidades e no contexto de um estudo.

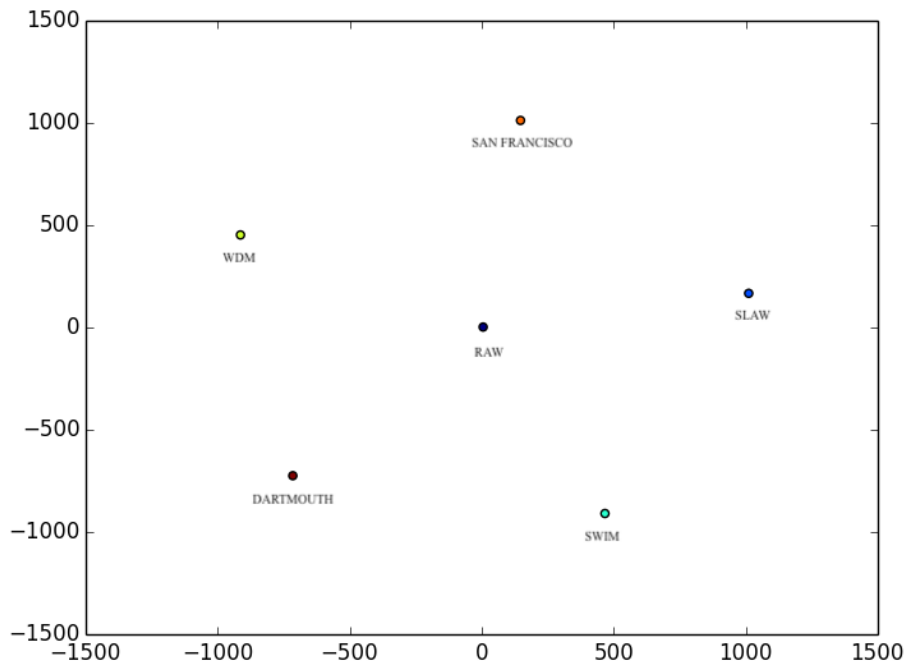


Figura 2. Resultado do T-SNE

Referências

- Akaike, H. (2011). *Akaike's Information Criterion*, pages 25–25. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Bai, F., Sadagopan, N., e Helmy, A. (2003). Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. In *INFOCOM 2003. Twenty-second annual joint conference of the IEEE computer and communications. IEEE societies*, volume 2, pages 825–835. IEEE.
- Baumann, R., Legendre, F., e Sommer, P. (2008). Generic mobility simulation framework (gmsf). In *Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 49–56. ACM.
- Bezerra, R. L., Campos, C. A. V., e de Moraes, L. F. M. (2009). Uma proposta de técnica para o ajuste de modelos de mobilidade em redes ad hoc e questionamentos sobre a adequação dos parâmetros envolvidos com base em dados reais. *Simpósio Brasileiro de Redes de Computadores (SBRC)(Brasil, 2009)*.
- de Melo, P. O. V., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Le Mouël, F., Loureiro, A. A., Addepalli, L., e Guangshuo, C. (2015). Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19–36.
- Dimatteo, S., Hui, P., Han, B., e Li, V. O. (2011). Cellular traffic offloading through wifi networks. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 192–201. IEEE.

- Ekman, F., Keränen, A., Karvo, J., e Ott, J. (2008a). Working day movement model. In *Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models*, MobilityModels '08, pages 33–40, New York, NY, USA. ACM.
- Ekman, F., Keränen, A., Karvo, J., e Ott, J. (2008b). Working day movement model. In *Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 33–40. ACM.
- Ferreira, F., Silva, T. H., e Loureiro, A. A. F. (2017). Slkit: An r package for property extraction and analysis of multiple sensing layers. In *Computer Networks and Distributed Systems (SBRC), 2017 XXXV Brazilian Symposium on*, pages 1184–1191. IEEE.
- Helgason, Ó., Kouyoumdjieva, S. T., e Karlsson, G. (2014). Opportunistic communication and human mobility. *IEEE Transactions on Mobile Computing*, 13(7):1597–1610.
- John Lu, Z. Q. (2010). The elements of statistical learning: Data mining, inference, and prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3):693–694.
- Karamshuk, D., Boldrini, C., Conti, M., e Passarella, A. (2011). Human mobility models for opportunistic networks. *IEEE Communications Magazine*, 49(12):157–165.
- Kim, M., Kotz, D., e Kim, S. (2006). Extracting a mobility model from real user traces. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications*, pages 1–13.
- Kosta, S., Mei, A., e Stefa, J. (2014). Large-scale synthetic social mobile networks with swim. *IEEE Transactions on Mobile Computing*, 13(1):116–129.
- Lee, K., Hong, S., Kim, S. J., Rhee, I., e Chong, S. (2009). Slaw: A new mobility model for human walks. In *IEEE INFOCOM 2009*, pages 855–863.
- Maaten, L. v. d. e Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605.
- Mei, A. e Stefa, J. (2009). Swim: A simple model to generate small mobile worlds. In *IEEE INFOCOM 2009*, pages 2106–2113.
- Munjal, A., Camp, T., e Navidi, W. C. (2010). Constructing rigorous manet simulation scenarios with realistic mobility. In *Wireless Conference (EW), 2010 European*, pages 817–824. IEEE.
- Piorkowski, M., Sarafijanovic-Djukic, N., e Grossglauser, M. (2009). CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from <https://crawdad.org/epfl/mobility/20090224>.
- Scholz, F. W. (2004). *Maximum Likelihood Estimation*. John Wiley Sons, Inc.
- Song, L. e Kotz, D. F. (2007). Evaluating opportunistic routing protocols with large realistic contact traces. In *Proceedings of the second ACM workshop on Challenged networks*, pages 35–42. ACM.
- Thakur, G. S. e Helmy, A. (2013). Cobra: A framework for the analysis of realistic mobility models. In *INFOCOM, 2013 Proceedings IEEE*, pages 3351–3356. IEEE.

Yang, D., Zhang, D., Chen, L., e Qu, B. (2015). Nazionelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications*, 55:170–180.