

移動データを用いた人間関係の抽出とその性質の分析

Retrieving and analyzing interpersonal relationship from large-scale mobility data.

浅谷 公威*1
Kimitaka Asatani

鳥海 不二夫*2
Fujio Toriumi

大知 正直
Masanao Ochi

森 純一郎
Junichiro Mori

坂田 一郎
Ichiro Sakata

This paper is extended abstract of our recently published paper[Asatani 18]. Interpersonal relationship affects our daily activity. In this paper, we consider the link between them by analyzing large-scale mobility data. In the past studies of human mobility, owing to the difficulty of associating mobility data with interpersonal relationship data, previous studies on the link between interpersonal relationships and mobility are limited to the specific activities of particular users. In this paper, we propose a method for detecting interpersonal relationships from mobility data, while distinguishing these relationships from those of familiar strangers such as commuters. In the method, persons who take diverse variations within the same activities are recognized as a pair. From IC card data covering the daily mobility of six million people over three years, we detected millions of frequently co-located pairs. Under certain conditions, most of the detected pairs are confirmed as not being familiar strangers, but rather to have an interpersonal relationship. Next, we analyzed the detected pairs and found that the density of the relationships between groups was divided by gender and age and was found to be asymmetric by gender. For example, an elderly male person is not likely to take trips as a pair with a same-gender elderly person, and this result is data-based evidence for the isolation of retired men. In addition, group trips are confirmed to have an extraordinary character and sometimes converge spatiotemporally. These findings indicate that interpersonal relationship is a strong factor to determine their mobility and group observation is potentially useful for event detection.

1. はじめに

本論文は [Asatani 18] で発表した内容の Extended abstract である。人の移動を理解し予測することは、感染症阻止、交通の最適化、イベント企画などの多くの応用につながる。これまでの研究の大部分は、人間の移動行動 [Gonzalez 08] や場所の特性 [Giannotti 11] に焦点をあてることで人間の移動を理解・予測してきた。いくつかの応用的な研究は、犯罪の可能性 [Bogomolov 14] と土地利用 [Grauwin 17] のパターンは人間の移動性に基づいて推定できることを示している。しかし、さまざまな動機から生じる人間の移動は完全には理解・予測することはできない [Song 10]。その大きな要因の一つは、人間の移動はしばしば人間関係によって動機付けられることであると考えられる。

ある先行研究はデータ解析により社会的相互作用が人間の移動の予測に重要な情報であることを示している [De Domenico 13]。社会的つながりと人間の移動性との関係についても解析が進んでいる。[Fan 17]。しかし、これらの研究の対象期間や対象となる人間は非常に限られている。そこで、データを用いない理論的なアプローチによる人間関係と移動についての考察も行われてきた [Miller 05]。このように人間の移動と人間関係に多くの注意が払われているにもかかわらず、一般の人間の日常におけるそれらに関する理解はあまり進んでいない。たとえば、どのような人々が一緒に移動しているか、人間関係が移動に与える影響は何か、グループでの移動の特性などの問いに対する考察は進んでいない。

より広い文脈で考えると、人間関係は人間社会の基本的要素であり、個人の精神的および肉体的健康と強く関連している [Newman 12]。これまで多くの研究者は、オンラインコミュニティや画像共有サービス [Cheung 15] における隠れた人間関係の推定や、人間関係とインターネット上での活動との関連を調査してきた [Pflieger 10]。ただし、これらのデータは各ユーザーの特定の活動に限定されており、日々の日常の行動がよく表されているデータを用いて人間関係を理解することが必要である。そのようなデータの一つとして、広範囲にわたって収集

された人間の移動データがあげられる。例えば、大都市に住む人々にとって、鉄道システムからの移動データは彼/彼女らの日常生活の活動の多くを反映している [Shimosaka 15]。したがって、人間関係データを人間関係と結び付けることが出来れば、私たちが日々の人間の相互作用をよりよく理解できるだろう。しかしながら、データの制限とプライバシーの問題から、このようなデータを組み合わせることには現実的に難しいと考えられる。

本研究では、人間の移動と人間関係を紐づけたデータがない状況で人間の移動と人間関係に関する上述の問を考察するために、人間の移動そのものから人間関係を推定する方法を開発した。データとして、関西圏(大阪、神戸、奈良、京都を含む日本で2番目に大きい都市圏)から600万人を対象として3年間に収集された22億の移動ログを用いた。電車による移動は、ICカードを使用して電車で移動する人々の日々の活動に不可欠であるため、これらのデータはさまざまな人間の日常的/非日常的移動を含む活動の多くを反映している。本研究では、一定数以上の出発/目的地の組み合わせにおいて同一トリップ(同じ目的地と同じ出発地をそれぞれ同時間)を行っている二人の人間の間に友人関係があるとみなした。

定義した手法を用いて移動データから頻繁に共存する数百万もの人間関係を抽出した。また、意図しない同一トリップを頻繁にする関係のない同乗者(通勤客など)と人間関係がある二人の組み合わせの間を区別するために、タイム時間ウィンドウ幅の変化に関して検出された二人の組み合わせの違いを観察した。ある条件下では、時間ウィンドウ幅の増加が検出された二人の組み合わせに与える影響が小さく、検出された二人の組み合わせの大部分は一緒に移動する強い意志を持つ人間関係を持っていると推測された。抽出した関係性を使い、大都市における我々は大規模な人間関係グラフを観察することにより、性別による人間関係の違いを明らかにした。さらに、誰かと一緒に移動することによる人間の行動の変化(例えば普段行かないところに行くなど)や、グループでの移動(人間関係がある人との同時移動)が時空間的に収束することを明らかにした。

連絡先: 浅谷公威, 東京大学工学系研究科, 東京都文京区本郷7-3-1, asatani@gmail.com

2. データ

分析に用いるデータは、2014年4月から2016年3月までの間の関西圏（大阪、神戸、奈良、京都を含む日本で2番目に大きな都市圏）における600万人の匿名ユーザーによる22億回の移動である。関西の人口の約25%にあたる500万人が3年間で100回以上鉄道を使用している。また、各移動は乗車駅と降車駅における乗車/降車の時間が分単位で記録されている。また、全体の33%のユーザーに関しては10年ごとの年齢ラベルと性別の情報が含まれている。また、データにはどのユーザー間に人間関係があるかを示す情報は含まれていない。

3. 手法

我々は人間関係を抽出するために異なる2つの方法を開発した。それぞれの手法はともに、頻繁に同時移動する二人の組合せを抽出している。1つ目は「同一トリップ検出」である。この方法では、少なくとも t_{id} の同一トリップをする二人の組合せが、頻繁に同時移動する二人の組合せとして検出される。時間ウィンドウ幅が $w = 1(\text{min})$ の場合、各個人が同じ目的地と同じ出発地をそれぞれ分単位で同時刻に入場/退場していれば同一トリップとなる。 $w \geq 2$ の場合、二人の個人の出発時刻と到着時刻の差が $w - 1(\text{min})$ 以内の場合に同一トリップと見なされる。少なくとも t_{id} の同一トリップを共有する人 x, y は頻繁に同時移動する二人の組合せとみなされる。1. この式の $IT^w(x, y)$ は、時間ウィンドウ幅 w でのユーザー x と y の間の同一の移動の有無 (1,0) を意味している。

$$P_{id}^{t_{id}, w}(x, y) = \begin{cases} 1 & n(IT_{(x,y), w}) \geq t_{id} \\ 0 & \text{else} \end{cases} \quad (1)$$

同一トリップ検出法では、同じ通勤ルートに共有するなどの見知らぬ同乗者が二人の組合せとして検出される可能性がある。仮に十分に大きい t_{id} の値をとったとしても、毎日同じルートの時間厳守の通勤者たちの間に人間関係があると判断してしまう。そのことを避けるため、十分に長期間にわたるデータがある場合に人間関係のある人々は多くの種類の移動と一緒にすると仮定し、少なくとも t_{ur} の異なる種類の目的地/出発地の組み合わせで同一トリップを行う二人の組合せを検出する「ユニークルート検出法」を提案した。より正式には、少なくとも t_{ur} の目的地/出発駅のセットで同一トリップを行うユーザー x, y は、人間関係がある二人の組合せと見なす。(式2を参照)。この式で、 $ITS_{(x,y), w}$ は、ユーザー x と y が時間ウィンドウ幅 w で同一トリップになる目的地と出発地の駅組 (順不同) を意味する。

$$P_{ur}^{t_{ur}, w}(x, y) = \begin{cases} 1 & n(\{ITS_{(x,y), w}\}) \geq t_{ur} \\ 0 & \text{else} \end{cases} \quad (2)$$

4. Results

4.1 人間関係の抽出

はじめに、頻繁に同時移動する二人の組合せの存在確率を考察する。図1(a)は、時間ウィンドウ幅 w を変化させながら、(a) 同一トリップ手法で t_{id} を変えながら検出された頻繁に同時移動する二人の組合せ数 N_{id} 、(b) ユニークルート手法で t_{ur} を変えながら検出された頻繁に同時移動する二人の組合せ数 N_{ur} を示している。図(a)に着目すると $w = 1$ では、 t_{id} を増やしたとき N_{id} にべき乗的に近く減少する。(a)の他の条件や(b)ではそれよりも若干急峻に減少する。一人一人がランダムに移動をすると仮定すると t_{id} や t_{ur} が大きくなったときの N_{ur} や N_{id} の減少は指数関数的であり、条件下の実測値のそれよりもはるかに緩く減少している。したがって、頻繁に同

時移動する二人の組合せが何かしらの要因で数多く存在していると考えられる。

頻繁に同時移動する二人の組合せが人間関係がある人同士なのか通勤などによる見知らぬ同乗者であるかを考察するために、検出された二人の組合せの数に対する時間ウィンドウ幅 w の影響を考察した。見知らぬ同乗者の二人の組合せは列車に乗ったり降車したりする時間お互いに調整して合わせることはないため、 w が大きくなるにつれて検出数は大幅に増加すると考えられる。一方、人間関係のある二人の組合せの検出数は w の変化に大きく影響を受けないと考えられる。彼/彼女らが同時に移動する場合、多くの場合はわずか数秒以内に改札を通過するためである。したがって、検出された二人の組合せから見知らぬ同乗者がすでに除外されている場合、時間ウィンドウ幅 w が大きくなっても検出される二人の組合せ数はそれほど増えないと考えられる。

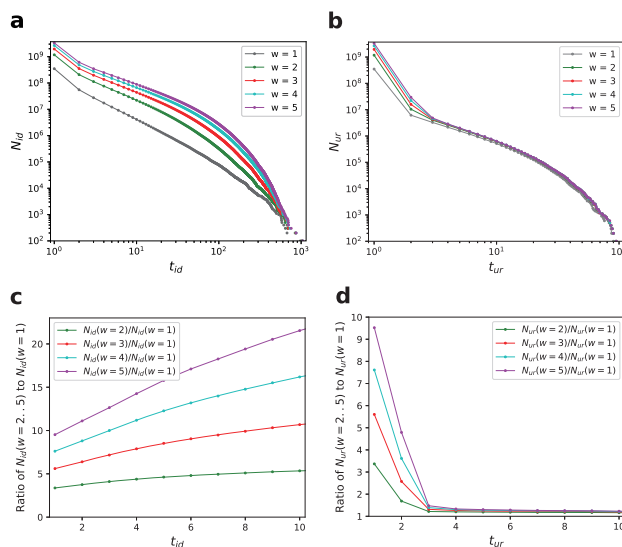


図1: (a, b) 同一トリップ検出法 (a) とユニークルート検出法 (b) で検出された、頻繁に同時移動する二人の組合せの数。横軸は、同一トリップ検出法の各しきい値 t_{id} , t_{ur} 、縦軸は頻繁に同時移動する二人の組合せの数を示す。各図では、時間ウィンドウ幅 w を1.5に変えた結果を異なる線でプロット。(c, d) $w = 2..5$ における頻繁に同時移動する二人の組合せの数の $w = 1$ の場合に対する割合

図1(a, c)より、同一トリップ手法では時間ウィンドウ幅 w の増加にともなって検出された頻繁に同時移動する二人の組合せ N_{id} の数が大幅に増加することがわかる。このことにより、同一のトリップ方法によって検出された検出された二人の組合せには、見知らぬ同乗者が多く含まれていると考えられる。その一方で、ユニークルート法の場合、 w の変化に伴う検出二人の組合せ数 N_{ur} の増加は、 $t_{ur} \geq 3$ ではそれほど大きくはない(図1(b, d))。この場合、検出された二人の組合せには見知らぬ同乗者がほとんど含まれていないと考えられる。言い換えれば、これらの検出された対の間の移動の時間差は短い時間に収束し、各人は正確に同じ時間に移動するという強い動機を持っていると推測される。

また、 $t_{ur} \geq 3$ においても $w = 2$ での二人の組合せ検出数が $w = 1$ よりも2割ほど大きい。しかし、 $w = 3, 4, 5$ と w を増やしたときの増加は少ない。この理由は、友人関係があるユーザーが同時移動するとき改札を通る時間が記録単位である分をまたがるものが一定の確率で存在するからだと考えられる。したがって、 $w = 1$ ではなく $w = 2$ を適切なパラメータと考える。

以上より、 $t_{ur} \geq 3$ のユニークルート法で頻繁に同時移動する二人の組合せのほとんどが、意図的に同一トリップをする人間関係を持っている二人の組合せだと推測できる。これらの理

由から、本論文ではパラメータ $t_{ur} = 3, w = 2$ を指定した一意のルート方法を使用する。

4.2 年齢性別による人間関係の密度

どのようなタイプの人々の間に人間関係が存在する確率が高いかを確認するために、グループ間の関係の相対密度 $\bar{D}(i, j)$ を図2にプロットした。 \bar{D} は全ユーザー間のリンク密度 D に対する、対象とするグループ間のグループ密度 $D(i, j)$ の割合である。

同年齢のグループの組み合わせの間には、多くの人間関係が存在する。10代から20代の人々は同じ性別のグループを形成する傾向があるが、30歳以上の人々は同じ年齢の男性/女性グループを形成する。この傾向は人間の生活を反映している：人は彼/彼女の友達と育ち、後でパートナーと家族を持っている。これらの結果は一般的に想像できる予想と一致している。ただし、性別および老齢学の観点から、人間関係密度の結果から考えられる重要な点は以下である。

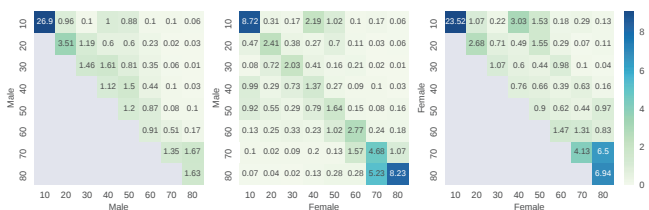


図2: \bar{D} : 性別・年齢別のグループ間の関係の相対密度 左から、男性/男性、男性/女性、女性/女性の年代別のグループの関係の相対密度 \bar{D} をプロットした。

- 40~50歳の年齢層の女性と10代の若者の人間関係の密度は、40~50歳の年齢層の男性と10代の若者のほぼ2倍である。これは、父親が母親ほど子供の世話をすることに関わっていないという事実を反映している。
- 高齢者(60歳以上)では、男性/男性の人間関係は、女性/男性および女性/女性の人間関係と比較してまれである。この傾向は、年配の男性の方が同行する友達が少ないという事実を反映している。
- 年配の男性と若い女性との間の人間関係は、若い男性と年配の女性との間よりも多く観察される。これは男性と女性との関係の非対称性を反映している。

パラメータ t_{ur} の値を2から10に変更しても同様の傾向の結果が得られた。これらの結果は、日本の日常生活の活動における著しい性差を強調しており、日本におけるジェンダーによる役割の差を示した研究や報告 [Schwab 16] を裏付けるものである。また、高齢男性が同性の友人と移動をすることはあまりないという結果は高齢男性の孤立の証拠である。これは、老齢学の専門家によるアンケート調査に基づく観察と一致している [Beach 14]。このように、大量の移動データを調査することによって、社会学および老年学の分野で提案された仮説を大規模なデータから確認したといえる。

4.3 グループ移動の特性

大都市における個人の移動は日々の日課により条件付けられている。[Wang 15, Gonzalez 08, Song 10]。ここでは、人間関係が人間の移動における他の決定要因であることを実証するために、はじめにグループ移動の非日常性を考察する。グループ移動とは人間関係が推測できる二人の組合せと一緒に移動した移動である。両側グループ移動とはグループ移動のうち乗車駅も降車駅が同じ移動、片側グループ移動とは乗車駅が降車駅のどちらかのみが同じ移動をさす。移動の戻り時間(すなわち、同じ駅への最初の到着時間とその次の到着時間との間の差)は、その移動の非日常性を性質を表す。移動の戻り時間が1日~数日と短ければ、その移動は日常的な移動と推測される。また、グループ移動におけるある駅への戻り時間の計測

には、対象者のその後の全移動を使用しており、グループ移動だけに限ってはいない。また、ユーザーが同じ日に駅へ何度か移動をする場合には、それらを単一の移動と見なす。戻り時間の(累積)確率分布は図3にプロットした。

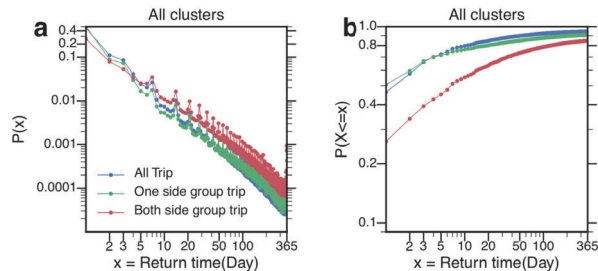


図3: (累積)戻り時間の確率分布。(a): 全移動の戻り時間の確率分布。(b): 全移動の累積確率分布。各図において、両側グループ移動(赤)、片側グループ移動(緑)、全移動(青)がプロットされている。

図3より、戻り時間の確率分布はべき乗則に従っていることが分かる。この結果は先行研究 [Song 10] の結果と一致している。べき乗則に比べ、7の倍数の日付と3日後の戻り時間の確率は高い値をとり、これは人間の週次の行動や通勤・通学における土日休みのスケジュールを反映しているといえる。

図3(a,b)、帰路時間の累積確率分布における全移動とグループ移動の違いを明確に示している。全体の移動の43%(図の青線3(b))では翌日に同じ駅に戻るのに対し、両側グループ移動はわずか26%(図の赤線)である。したがって、両側のグループ移動非日常のものであると考察される。この結果は人間関係が非日常的な移動をもたらすことを示している。

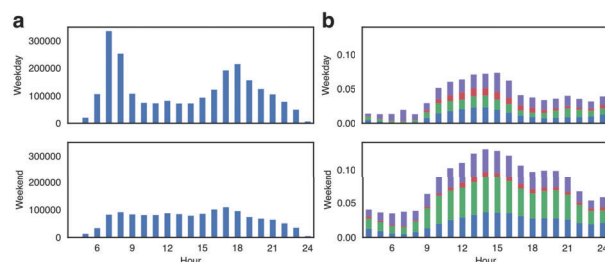


図4: (a) 時間別移動数平日と週末(休日を含む)で計算します (b) グループ移動の時間比率各図で、平日/週末の値を上下にプロット。

人間関係のある人々との同時移動が非日常的な移動である確率が高いことの要因を考察するために、いつ、どのような場所へ知り合いとで一緒に移動するのかを調べた。ユニークルート手法では、全移動の9.7%がグループ移動として識別され、これらは特定の時間に発生する傾向があった。図4(b)に示すように、両側のグループ移動の割合は平日の日中と週末に増加した。これららの時間には一人での移動が多い通勤が少ないためだと考えられる。また、グループ移動の平均時間は25分程と、全移動の平均時間の28分より短い結果となった。これは、グループ/二人の組合せの意思決定プロセスの中で、人々が長旅のための高コストを避けていることを示唆している。

4.4 グループ移動の時間空間的な集中

以上の分析でグループ移動は非日常的な性質を持っていることが分かった。ここでは、グループ移動先が特定の時刻に特定の駅に時間空間的に集中していることを示す。大きな競技場の近くの駅(図5(a))と他の駅(図5(b,c))のグループ移動数と全移動数の時系列を観察した。図5(a)では、いくつかの明確なピークが見て取れるが、これらはスタジアムでイベントが開催された日とほぼ正確に一致している。一方で、1日の総移動数のピークはイベントが発生した人との相関は比較的少ない。

また、都市の住宅街駅 (b,c) ではグループ移動数を観察することによって、2月上旬 (b) と12月中旬 (c) にそれぞれイベントが発生したことがわかる。

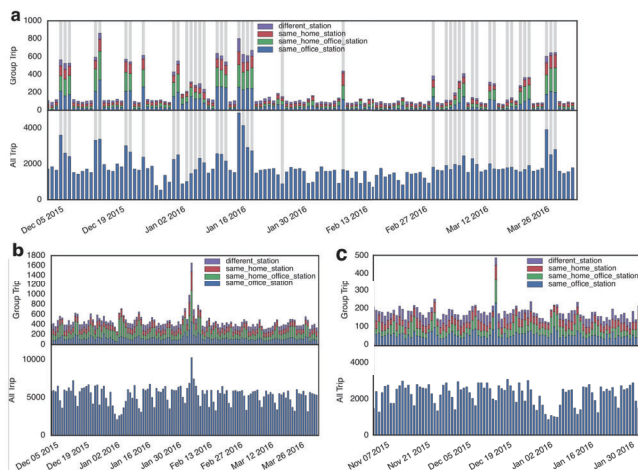


図 5: 1日のグループ移動数と総移動数。1日のグループ移動数 (上) と総移動数 (下) をプロット。(a) イベント会場の近くの駅。実際にイベントがひらかれた日を灰色でハイライトしている。(b, c) 住宅地にある駅。

これらの観察から、モビリティを介した人間の相互作用の一時的な収束が存在する。したがって、友達との移動はイベント検出の信頼できる指標である。つまり、ある駅における友人と一緒に降りる人の割合は、駅の近くでイベントが発生していることを示す一つの手法として有効である。

5. 考察

本稿では、大規模モビリティデータから人間関係を抽出する手法を提案した。提案手法において多様な種類の同一移動をとる人々が二人の組合せとして検出される。実際のデータを分析したところ、検出された二人の組合せには見知らぬ同乗者はほとんど含まれていないと推定された。このアプローチは人間の相互作用をデータから推測することに応用可能だと考えらる。鉄道移動データに加えて、記録される情報の量が増加すると、潜在的に人間の相互作用に関する情報、例えばGPSログデータ、ウェブサービスの活動や購入ログが蓄積されている。いくつかの研究では、ソーシャルネットワークや画像共有サービスに基づいて、ユーザー間の関係を推定しようとしている [Cheung 15]。本研究で提案した、同じ活動グループの人々を誤って含まずに友人を検出する方法は、これらのデータが十分に長い期間にわたって記録されていれば、他のデータセットにも適用可能と考えられる。

以前のモビリティに関する研究では、人間-場所の関係性を考慮することにより人間の移動を理解・予測してきた [Wang 15]。視点を変えて考えると、多くの移動は友人や会社・学校の人と合うことを目的として行われるので、人間の移動は人間-人間の間をベースに理解することも可能である。本稿では、人間関係がモビリティのもう一つの決定要因であることを確認した。我々の結果が人間の移動性のある基本的な側面を明らかにするためのエージェントベースのアプローチに貢献することを期待している。

さらに、移動から推測された人間関係の密度は性別や年齢によって異なり、性差による日常生活の違いがデータから読み取れた。例えば、年配の女性と比較して、年配の男性が互いに交流することはめったにない。年配の男性の孤立は私たちの社会の中で重要な課題である [Beach 14]。これに加えて、30~50歳の女性と若者の間に存在する関係の確率は、30~50歳の男性と若者のそれとは有意に異なる。中年の男性が中年の女性より若い人と移動する可能性が低いという結果は、子育てにおける性差を表している。この結果は、政策決定やイベント計画に

適用可能な証拠となりうるだろう。

参考文献

- [Asatani 18] Asatani, K., Toriumi, F., Mori, J., Ochi, M., and Sakata, I.: Detecting interpersonal relationships in large-scale railway trip data, *Journal of Computational Social Science*, Vol. 1, No. 2, pp. 313–326 (2018)
- [Beach 14] Beach, B. and Bamford, S.: Isolation: the emerging crisis for older men, *London: Independent Age* (2014)
- [Bogomolov 14] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., and Pentland, A.: Once upon a crime: towards crime prediction from demographics and mobile data, in *ACM ICMI*, pp. 427–434 ACM (2014)
- [Cheung 15] Cheung, M., She, J., and Jie, Z.: Connection discovery using big data of user-shared images in social media, *IEEE Transactions on Multimedia*, Vol. 17, No. 9 (2015)
- [De Domenico 13] De Domenico, M., Lima, A., and Mu-solesi, M.: Interdependence and predictability of human mobility and social interactions, *Pervasive and Mobile Computing*, Vol. 9, No. 6, pp. 798–807 (2013)
- [Fan 17] Fan, C., Liu, Y., Huang, J., Rong, Z., and Zhou, T.: Correlation between social proximity and mobility similarity, *Scientific Reports*, Vol. 7, (2017)
- [Giannotti 11] Giannotti, F. and Nanni, M. e.: Unveiling the complexity of human mobility by querying and mining massive trajectory data, *VLDB*, Vol. 20, No. 5, pp. 695–719 (2011)
- [Gonzalez 08] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L.: Understanding individual human mobility patterns, *nature*, Vol. 453, No. 7196, p. 779 (2008)
- [Grauwin 17] Grauwin, S. and Szell, M. e.: Identifying and modeling the structural discontinuities of human interactions, *Scientific Reports*, Vol. 7, (2017)
- [Miller 05] Miller, H. J.: Necessary space–time conditions for human interaction, *Environment and Planning B: Planning and Design*, Vol. 32, No. 3, pp. 381–401 (2005)
- [Newman 12] Newman, P. and Matan, A.: Human mobility and human health, *Current Opinion in Environmental Sustainability*, Vol. 4, No. 4, pp. 420–426 (2012)
- [Pflieger 10] Pflieger, G., Rozenblat, C., Mok, D., Wellman, B., and Carrasco, J.: Does distance matter in the age of the Internet?, *Urban Studies*, Vol. 47, No. 13, pp. 2747–2783 (2010)
- [Schwab 16] Schwab, K., Samans, R., and Zahidi, e.: The Global Gender Gap Report 2016, in *World Economic Forum* (2016)
- [Shimosaka 15] Shimosaka, M., Maeda, K., Tsukiji, T., and Tsubouchi, K.: Forecasting urban dynamics with mobility logs by bilinear Poisson regression, in *Proceedings of the UbiComp 2015*, pp. 535–546 ACM (2015)
- [Song 10] Song, C., Qu, Z., Blumm, N., and Barabási, A.-L.: Limits of predictability in human mobility, *Science*, Vol. 327, No. 5968, pp. 1018–1021 (2010)
- [Wang 15] Wang, Y. and Yuan, N. J. e.: Regularity and conformity: Location prediction using heterogeneous mobility data, in *SIGKDD*, pp. 1275–1284 ACM (2015)