Institut für Parallele und Verteilte Systeme
Abteilung Verteilte Systeme
Universität Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Diplomarbeit Nr. 2903

A Mobility Model for the Realistic Simulation of Social Context

Daniel Fischer

Studiengang:	Informatik
Prüfer:	Prof. Dr. rer. nat. Dr. h. c. Kurt Rothermel
Betreuer:	Dr. Ing. Klaus Herrmann
begonnen am:	29. Januar 2009
beendet am:	31. Juli 2009
CR-Klassifikation:	C.2.1, C.4, I.6.0

Abstract

The widespread use of user-carried devices with short-range communication leads to networks characterized by high dynamics, sporadic connectivity, and strong partitioning. In such networks, connectivity between mobile nodes is strongly influenced by sociological aspects. To enable the evaluation of mobile applications which communicate in such networks, we require an appropriate mobility model.

In this thesis, we have designed and implemented a mobility model which focuses on the simulation of social context. It takes an arbitrary weighted social network as input and reflects its structural properties in its mobility scheme. Based on this approach, our model allows to integrate recent advances in the research of complex social networks. In addition, we focus on the simulation of different typical human characteristics such as the periodical reappearance at preferred locations and movement in groups. Furthermore, our model allows the integration of mobility models which concentrate on geographical aspects such as modeling obstacles or realistic movement between locations.

We provide experimental results that show that our model reflects the input social network with an accuracy of up to 99%. In addition, we show that our model captures the characteristics measured in traces of human mobility, which shows the validity of our approach. The generalizational character of our model enables the fast integration of future research results in the areas of human mobility and complex social networks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In recent years, powerful and yet wearable devices with wireless network interfaces have become affordable and are already in widespread use among the populace. By utilizing only wireless short range communication, a disconnected adhoc network emerges. Connectivity between mobile nodes is created only sporadically dependent on the movement of the users wearing such devices. However, human movement is strongly influenced by social relations. We do not meet other people on a random basis but meet some people more frequently and regularly, e.g. friends, colleagues, family, or more general, people who are part of our **social context**.

In such a very dynamic network, end-to-end connectivity between two arbitrary nodes cannot be assumed. Nevertheless, communication is still possible, using a store-and-forward approach. The disadvantage of this approach is that significant delays are introduced. However, not all mobile applications depend on the real time delivery of messages. A prominent application of this class is the e-mail service.

One may argue that infrastructure could be used for a faster communication such as GSM cell towers for mobile phones. Yet, in many cases, the use of such an infrastructure involves costs. In other cases, infrastructure may not be available at all.

Mobile applications and distributed algorithms which run on wireless wearable devices may exploit knowledge about the social context of its carrier, e.g. for forwarding decisions if a message must be transmitted between two arbitrary mobile nodes. Consider the scenario that we want to deliver a message from user A's device to user B's device and know that another user C has a social relation with both A and B. In this case, the device of C may be a good candidate as the next hop for the message since C meets frequently with both users.

As we will discuss in more detail over the course of this thesis, social networks, i.e. networks of users and their relations, have characteristic structural properties. Because the social network directly influences the mobility of nodes, mobile applications (which includes protocols and distributed algorithms) may exploit the structural properties of a social network. For example, some devices carried by users characterized by many social relations may be utilized for a fast information dissemination among the participants.

To enable the research of mobile applications which exploit knowledge about the social context of its carrier or exploit structural properties of the social network, a thorough evaluation is necessary. An evaluation in real environments based on human-carried devices is very elaborate and costly. Hence, the evaluation of such mobile applications depends in most cases on simulations. In this thesis, we propose a discrete event simulator based on a mobility model which simulates the movement of human-carried devices influenced by their social context, i.e. a *social* mobility model.

1.2 Goals and Focus of the Thesis

The goals behind the design of our model are the following:

- It is important that a mobility model is as realistic as possible. Thus, it should capture characteristics observed in real measurements of human mobility. This avoids a subjective evaluation of realism.
- Many mobility models focus on *micro-mobility* aspects such as speed, direction, or restriction through obstacles, i.e. geographical concepts. Other models focus on the simulation of sociolocial aspects. However, so far, both strands are separate. To provide for more realism and to adapt the simulation to specific target scenarios, it is imperative to integrate geographical concepts into a social mobility model. This thesis represents the first steps towards this integration.
- Different existing (social) mobility models typically are coined by a fixed scale of space and time. In addition, they focus on the simulation of specific aspects of human behavior. To avoid that researchers have to employ several different mobility models, we advocate that existing approaches should be generalized. The researcher may then change the behavior of the model by simply adjusting the parameters.
- The structure of social networks has been intensively researched in the last ten years. However, many different models exist and only few results about the mapping between the social network and the mobility of users are known. Thus, a mobility model should be independent of specific assumptions about the structure of social networks. Instead, we focus on enabling the integration of results in this research area into our mobility model.

We separate our approach from other approaches which focus on creating connectivity between nodes on the individual *user level* such as shared agendas [ZHL06, EKKO08] or common membership in a social community [GGP08]. We argue that it is questionable whether such properties result in the realistic reflection of the structural properties of social networks. Our approach behind incorporating social context is to use a specific weighted social network as a basis to create our mobility scheme. Hence, we create connectivity between mobile nodes on the *social network level*. This social network is provided as the primary input of the model. We *reflect* this network by creating meetings between simulated mobile nodes whose users share a social relation. By generating connectivity based on an input social network, possibly generated from well researched models or extracted from real data, we avoid making decisions on the user-level. Instead, we separate such decisions from our model.

1.3 Outline

The remainder of this thesis is organized as follows. Chapter 2 gives some background information on mobility models and the research results on the analysis of social networks. We also discuss mobile applications which may benefit from our model. In Chapter 3, we analyze and identify important characteristics of human mobility that should be captured by a realistic social mobility model. In Chapter 4, we critically review existing social mobility models based on the identified requirements and point out possible improvements.

In the main part of this thesis, we propose our social mobility model (Chapter 5) and its prototypical implementation (Chapter 6). Subsequently, we validate our model based on the identified requirements (Chapter 7). During the evaluation, we compare characteristics exhibited by our model to the characteristics of real measurements of human mobility. To finish the main part of this thesis, we sketch some possible improvements to obtain further realism (Chapter 8).

To conclude this thesis, in Chapter 9, we first summarize our work and point out its contributions. Additionally, we give a critical review of the results and an outlook on possible further work.

Chapter 2

Background

After the introduction of some basic definitions, we provide an overview of the structural properties discovered in social networks and introduce corresponding models. Then, we look at mobile applications which may utilize those properties and may benefit from our proposed model. Furthermore, we give some background information on mobility models and argue why we require them for the evaluation of the discussed applications.

2.1 Definitions and Terminology

In this section, we give some basic definitions and discuss a number of terms which are frequently used over the course of this thesis.

By a **social relation** between two individuals, we refer to an arbitrary type of relationship, like partner, family, friend, colleague, or client. A social relation may be measured by a single numeric weight, which represents a measure of the strength of the relationship.

A **social network** consists of a number of *actors* and their social relations. It can be represented by a graph or an adjacency matrix, known in sociology as a *sociogram* [Sco00]. Formally, we use a graph representation where the actors are represented by nodes and their social relations are represented by edges. This yields the following definition:

Definition 1 (Social Network) *A social network is an undirected graph* $G = (V, E, w)$ *where* $V = \{v_1, .., v_n\}$ *denotes the set of nodes or actors, and* $E \subseteq \{\{x, y\} \mid x, y \in V\}$ *defines the social relations with* $\{u, v\} \in E$ *if nodes* u *and* v *share a social relation. Each pair of nodes may be annotated with a weight* $w : V \times V \to [0, 1]$ *with* $w(u, v) > 0$ *if* $\{u, v\} \in E$ *and* $w(u, v) = 0$ *else.*

In particular, if two nodes do *not* share a social relation, their corresponding weight is zero. We use the same definition without the weight function w for a general (not necessarily social) network. We call a node $v \in V$ a **social acquaintance** of $u \in V$ if $\{u, v\} \in E$. This is equivalent with the definition of 'neighbor', used for general networks. The distance $d(v, u)$ between two nodes $u, v \in V$ is defined as the length of the shortest path (i.e. the geodesic distance) between u and v.

The term 'social context' is used very ambiguously among different researchers. Adams et al. include three components in their definition of social context: *Locations of significance*, *regularities of the user behavior*, and *social ties* [APV08]. Smith et al. differentiate between the personal social context, which includes the user's friends and communities, and the community social context, which includes the user's role and identity in different communities [SBGL08]. De Choudhury et al. refer by social context to the 'degree of overlap of friends between two people' and their 'patterns of participation in communication' [DCSJS08].

Other authors do not provide a definition of social context at all [MT07, SSX07, KJRN05], but simply use the term, implicitly assuming that the reader has the same notion on 'social context'. Based on the use of the term, the authors typically refer either to the social acquaintances of a user in general, or only to social acquaintances which are physically co-located with the user. Note that most authors include many different concepts in their definition of social context. We advocate a more narrow and precise definition and the employment of other terms for different concepts. Furthermore, we note that community information and overlap of friends is implicitly encoded in the social network. Thus, we introduce the following informal definition of social context:

Definition 2 (Social Context) *The social context of an individual i defines i's local view of the social network. This includes i's immediate social acquaintances and possibly incomplete information about the structure of the social network within a distance of $d > 1$ around i.*

We do not only include the immediate social acquaintances in our definition since additional information may be exploited by mobile applications, e.g. knowledge about the friend of a friend.

In the remaining thesis, we use the term 'mobile node' or 'node' both for mobile wireless devices *and* the user who carries the device. The individual meaning may be recognized by the context. Furthermore, we use the term 'mobile applications' to refer to all kinds of software that runs on human-carried devices. This includes protocols and algorithms of arbitrary network layers.

2.2 The Structure of Social Networks

In the last decade, the study of complex networks [BLM+06, New03] has received great attention. Complex networks are *large-scale* networks with non-trivial statistical properties compared to the well-known random graph model. In this section, we focus on recent advances in the research of social networks, which belong to the class of complex networks. In particular, after defining some basic metrics on graphs, important structural properties of social networks are discussed and compared to the properties of random graphs. Subsequently, we introduce social network models which try to capture the discussed properties. We will later use these models to generate input networks for the implementation of our proposed social mobility model. This section also shows the variety of network models existing today and, thus, further motivates the choice of using a network as input of our model.

2.2.1 Basic Graph Metrics

The following metrics, among others, are used to characterize the structural properties of complex networks. The average length of all shortest paths in a graph is denoted as the **average path length** $L(G)$ [New03]:

$$L(G) = \frac{1}{n(n-1)} \sum_{v,u \in V} d(v,u).$$

From a social network perspective, the clustering coefficient is a measure for the probability that two individuals which share a common social acquaintance know each other [New03], i.e. the probability that the statement

$$\{u,v\} \in E \wedge \{u,w\} \in E \Rightarrow \{v,w\} \in E$$

holds. Several definitions for the clustering coefficient exist in the literature. A widely used definition for the local clustering coefficient of a single node $v \in V$ is the following [BLM+06]:

$$C(v) = \frac{2\,|\{\{u,w\} \in E \mid \{v,u\},\{v,w\} \in E\}|}{k(v)(k(v)-1)},$$

where $k(v)$ denotes the **degree of** v, i.e. the number of neighbors. In other words, $C(v)$ represents the ratio between the actual and the maximum possible number of edges between all neighbors of v. The clustering coefficient of a network G is then defined as the mean clustering coefficient of all nodes:

$$C(G) = \frac{1}{n} \sum_{v \in V} C(v).$$

2.2.2 Properties of Random Graphs

As will become apparent in the following sections, the properties of social networks differ compared to the well known random graphs. A common model for random graphs is the Erdös-Renyi model [ER59]. An instance $G_R(n,p)$ of this model represents a graph of n nodes with a probability of p that an edge exists between two arbitrary nodes. The node degree distribution, i.e. the probability that an arbitrary node has the degree k, is well approximated by the Poisson distribution

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where λ denotes the expected node degree $\lambda = pn$. If $p > \frac{ln(n)}{n}$, the graph is expected to be totally connected. The average path length is asymptotic to

$$L(G_R(n,p)) \sim \frac{ln(n)}{ln(\lambda)}$$

and thus logarithmic for a fixed p. The mean clustering coefficient of Erdös-Renyi graphs corresponds to $C(G_R(n,p)) = \lambda/n = p$.

2.2.3 Properties of Social Networks

The following structural properties have been discovered in social networks.

Small World Property

Networks characterized by a small average path length with respect to their size are known as 'small world' networks. The 'Small World Phenomenon' was originally discovered by Pool and Kochen in the 1950ies, but was not published until two decades later [PK78]. In the meanwhile, Milgram learned about the work of Pool and Kochen and conducted a today famous experiment [Mil67]. He gave 296 individuals a letter with the recipient being a certain stock broker living in Boston. He told them to forward the letter to acquaintances who they knew on a first name basis and who would be (presumably) closer to the final recipient. The immediate recipients forwarded the letter again until the letter reached the final recipient. In every forwarding step, the forwarders wrote down their names. About 20% of the letters reached the final recipient. However, only an average of 5.5 forwarders were required. This was a surprising result and Milgram concluded that the people in the U.S. are only separated by about 6 links of acquaintanceship.

Milgrams result was later confirmed by many other experiments on social networks, for example in networks of scientific collaboration [New01] and e-mail networks [DMW03]. The small world property can also be observed in real traces of human mobility [HH06a].

High Clustering Coefficient

Social networks typically exhibit a high clustering coefficient compared to random graphs with an equal number of nodes and edges [WS98]. This property of social networks is actually not surprising. It may be explained by the theory of *triadic closure* [Rap57, Gra73], which originates from the social sciences: individuals having two social acquaintances who do not know each other are compelled to close the triangle, for example, by introducing the social acquaintances to each other.

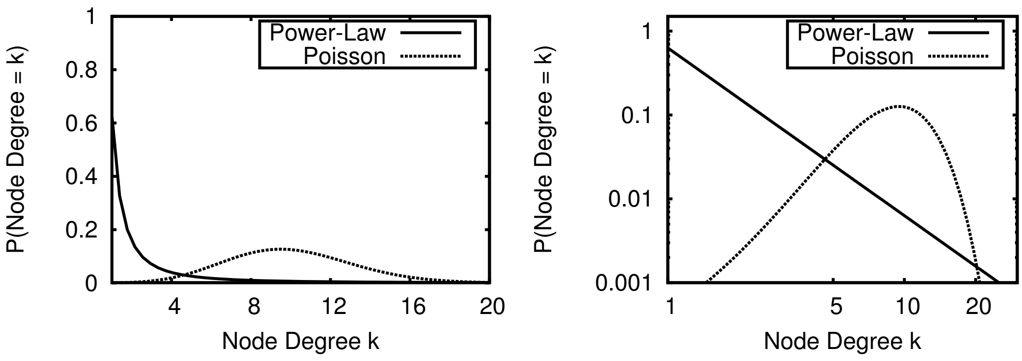
Power-law Node Degree Distribution

A property frequently observed in social networks is the power-law distribution of the node degree [BA99]. A power-law distribution can be approximated by

$$P(k) \sim k^{-\lambda} \text{ with } \lambda > 0.$$

Networks characterized by a power-law node degree distribution are called *scale-free* networks. In such networks, the majority of nodes have a small node degree. However, a small number of nodes called *hubs* have a large number of edges. Hubs play an important role because they connect different groups of nodes. This characteristic may be utilized. For example, Adamic et al. describe methods to exploit the power-law degree distribution to efficiently search in scale-free networks [ALPH01]. Figure 2.1 shows the node degree distribution of scale-free networks (solid) in comparison to the node degree distribution of random graphs (dashed). Note that a power-law corresponds to a straight line if a log-log scale is used.

The class of scale-free networks contains (among others) the social networks of movie-actors [BA99], sexual contacts [LEA^{+}01], phone-calls [ACL00], and science collaboration [BJN^{+}02]. Albert et al. [AB02] describe further examples.



(a) ***Uniform Scale.*** **(b)** ***Log-Log Scale.***

Figure 2.1: Power-law and Poisson node degree distribution.

Though many social networks are scale-free networks, this is not always the case. An example of this class is the network of Fortune 1000 (the thousand companies with the highest revenues in the U.S.) company directors of 1999 [DYB03]. In this network, two directors share a social relation if they sat on the same board together in the year 1999. Thus, the degree of a node measures the number of others with whom a director sat on boards. The node degree distribution of this network follows an approximately exponential tail [NWS02]. Similar results have been observed during the investigation of a friendship network of 417 high school students [FS64]. Another class of social networks is characterized by a power-law distribution followed by a sharp cut-off [ASBS00].

The literature explains the difference between social networks with and without the scale-free property due to the presence of 'costs' associated with the creation of social relations [JGN01, ASBS00, NWS02]. In the network of movie-actors, an edge between two nodes is created by simply starring *once* in a movie together or in the telephone network by making *a single* phone call. There is no need for further action to maintain a created edge in such networks. However, it involves time to maintain a friendship or to work on a company board. Since the available time of the actors is limited, there exists an upper bound on the number of edges one person can actively 'maintain'. This results in the typical sharp cut-off found in several social networks. Results from Hill and Dunbar [HD03], which show that a typical human is only able to maintain about 150 relationships (Dunbar's Number) based on the size of their neocortex, confirm these results.

Assortative Mixing

Newman reports that a variety of social networks are *assortative mixed* [New02]. This property describes the preference for high-degree nodes (hubs) to be connected to other high-degree nodes. Newman also investigates technical and biological networks and observes that this property does not hold for them. He states that networks with assortative mixing percolate better and are more robust against the removal of nodes. Thus, he argues that models should capture assortative mixing.

Community Structure

Community structure is another property of social networks [GN02] [GDDG+03]. A community can informally be described as a highly connected subgraph. Communities in social networks are hierarchical [ZSHD05]. According to Newman and Park, community structure can explain the strong clustering and assortative mixing observed in social networks [NP03]. Great effort is put in the research of community detection algorithms [New06, RCC+04, DDGDA05]. Recent work focuses on the detection of *overlapping* communities [WLZ08, PDFV05].

2.2.4 Models for Social Networks

Several models have been proposed which try to capture different properties of social networks.

Caveman Model

A very simple model for social networks has been proposed by Watts [Wat99], called the caveman model. Initially, nodes are partitioned into disjoint, fully connected subgraphs (caves). Because the subgraphs are completely isolated, Watts compares them to primitive humans in caves. Subsequently, each node is rewired to a node in another cave with probability p. Figure 2.2 shows three initial caves (left) and the result after the rewiring process (right).

Obviously, this model creates a strong community structure. p controls the degree of isolation between two caves. Because of the initial fully connected subgraphs, the model results in a high clustering coefficient. However, the average path length is comparable with that of a lattice i.e. linear in the number of nodes. Hence, the caveman model does not resemble a small world.

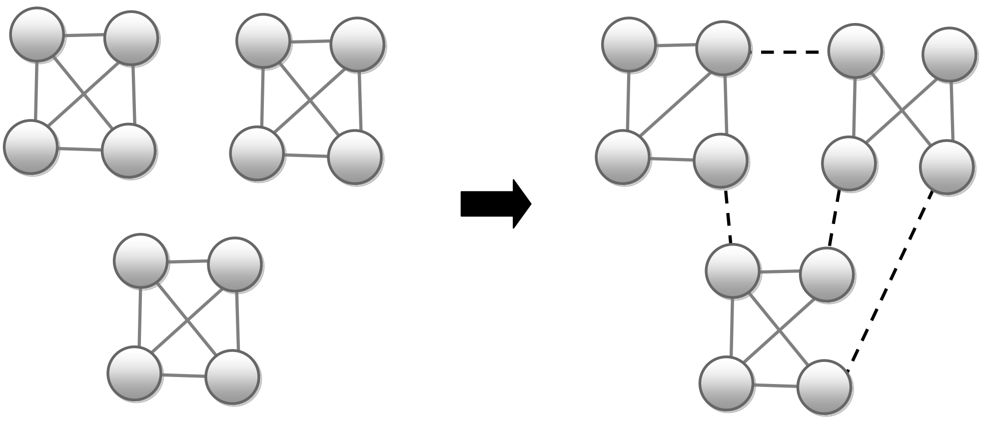


Figure 2.2: Example for the rewiring process in the caveman model.

Barabasi-Albert Model

In 1999, Barabasi and Albert [BA99] proposed a model for *growing* scale-free networks. The basic principle is the combination of two mechanisms:

1. *Growing Network.* The network is initialized with a small number m_0 of nodes, for example created by the random graph model. For every subsequent step, a node joins the graph and attaches (i.e. creates edges) to $m \leq m_0$ existing nodes.
2. *Preferential attachment.* Newly added nodes attach preferably to nodes which already have a high degree. More precisely, the probability to create an edge to a node v is defined by

$$P(\text{attach to node } v) = \frac{k(v)}{\sum_{u \in V} k(u)}$$

In other words, the probability for an attachment to v is proportional to the degree of v (the rich get richer). This leads to a power-law distribution of the node degree with $\lambda \approx 3$. Subsequent investigations show that the average path length approximates $L \sim \frac{\log n}{\log \log n}$ [AB02]. Thus, the Barabasi-Albert model satisfies the small world property. However, the clustering coefficient follows a power law $C(G) \sim n^{-0.75}$ and therefore decays towards zero for $n \to \infty$. This is considered to be unrealistic for many networks, including social networks. The following model addresses this issue.

Holme-Kim Model

The Holme-Kim (HK) model [HK02] aims to produce networks which are characterized by a scale-free node degree distribution *and* a high clustering coefficient. It is based on the Barabasi-Albert model. The basic idea is that nodes not only attach by preferential attachment (PA) but also attach to neighbors of already connected nodes, which creates a closed triad. The latter is called a *triad formation* (TF) step. More precisley, if a node joins the network and should attach to m nodes, a PA step is performed initially. For the next $m - 1$ attachments, either a TF or a PA step is performed with probability p_t and $1 - p_t$, respectively.

Figure 2.3 shows an example with $m = 5$. Node v joins the network and performs two PA steps (thick dashed lines) selecting nodes i and j. Subsequently, three TF steps are performed, which leads to the additional attachment to three different neighbors (thin dashed lines) of i and j.

The degree distribution of the HK model yields a power-law distribution $P(k) \sim k^{-3}$. The parameter p_t controls the strength of the clustering up to a clustering coefficient of $C(G) = 0.5$ for $n \to \infty$. The average path length scales logarithmic, similar compared to the Barabasi-Albert model.

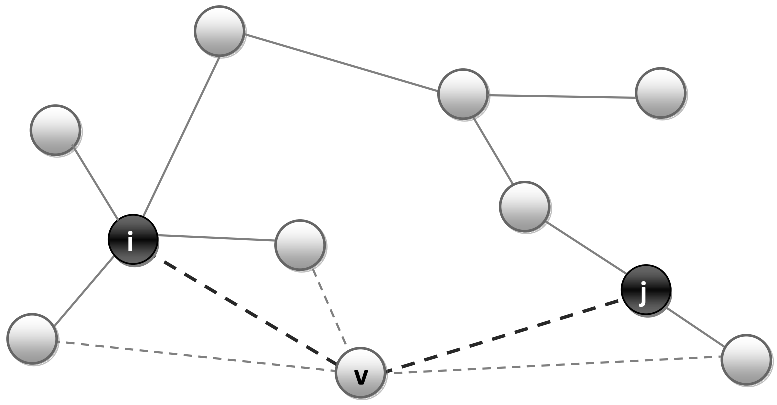


Figure 2.3: Example for two preferential attachment steps (thick dashed lines) combined with three triad formation steps (thin dashed lines).

Toivonen Model

The model proposed by Toivonen et al. [TOS+06] bears some similarity to the HK model. The main difference is that instead of preferential attachment, a random attachment is employed. An instance of the Toivonen model is constructed as follows:

1. Start with a small network of m_0 nodes (similar to the BA/HK model)
2. Attach to an average of $m_r \geq 1$ random nodes (initial contacts)
3. Attach to an average of $m_s \geq 0$ neighbours of nodes selected in the previous step (secondary contacts)
4. Repeat steps 2 and 3 to reach the desired network size

Roughly speaking, the initial contacts create bridges between communities, and the secondary contacts strengthen the community structure. According to the authors, it is important that m_r and m_s are not constant for each iteration (in constrast to the BA/HK model) to obtain the desired properties. The model shows strong assortative mixing, significant clustering, and a logarithmic path length. The degree distribution can be approximated by a power-law followed by an exponential decay. The authors have compared the community structure to real world social networks and show a strong match in terms of community size distribution.

Discussion

As shown, different social network models capture different properties. Because properties of social networks differ to some degree, in particular in terms of the scale-free property, it is not possible to identify the 'best' social network model. We summarize the properties of the discussed social network models in Table 2.1.

Model	**Log. L(G)**	**High C(G)**	**Scale-free**	**Comm.**	**Ass. Mixed**
Random	Yes	No	No	No	No
Caveman	No	Yes	No	Yes	No
BA	Yes	No	Yes	No	No
HK	Yes	Yes	Yes	No	No
Toivonen	Yes	Yes	with cut-off	Yes	Yes

Table 2.1: Properties captured by the discussed network models [New02, TOS$^+$06, BLM$^+$06].

2.3 Mobile Applications using Social Context

In Chapter 1, we have outlined a system model for mobile applications which may benefit from the evaluation based on a social mobility model. The described scenario falls into several research areas:

- **Opportunistic Networks** [PPC06] do not assume any complete end-to-end path between two nodes. Instead, connectivity is utilized in an opportunistic manner.
- **Delay tolerant networks (DTNs)** [Fal03], in contrast to opportunistic networks, assume existing connected regions (e.g. groups of people). Mobile nodes provide periodic connectivity between regions. It can be argued that the class of opportunistic networks contains the class of DTNs. However, the terms 'opportunistic network' and 'delay tolerant network' are often used interchangeably [PPC06].
- **Pocket Switched Networks (PSNs)** [HCS$^+$05] fall under the DTN space. PSNs aim to utilize the resources of heterogeneous devices like laptops, PDAs, or mobile phones transparently. Such devices may be carried in the *pockets* of people, hence the name. The research in this area concentrates specifically on the problems related to this heterogeneity.

In such networks, communication between two arbitrary nodes is typically conducted by buffering messages and waiting for connectivity opportunities. Of course, this comes at the price of an increased delay. Most work in the described areas focuses on opportunistic routing [MGC$^+$07, SNT08, SHR08, LDS04]. Because of sparse connectivity, efficient routing poses a challenging problem. Many protocols focus on epidemic routing, which leads to an optimal delivery delay. However, since a message is flooded over the network, such a strategy is very expensive and not scalable. Thus, efficient protocols must forward data by employing only a small subset of nodes as forwarders. Random selection of nodes leads to huge delays. Thus, possible forwarders must be selected according to some other information, e.g. statistical data or knowledge about social context.

In the following, we give some examples of mobile applications which may benefit from a social mobility model. Many focus on opportunistic routing, but other classes of applications are introduced as well. They are all based on a system model similar to the above described research areas.

2.3.1 Exploiting the Structure of Social Networks

Daly and Haahr propose a technique for routing in DTNs [DH07]. This technique exploits, among others, the small world property of social networks by using two metrics known from the research on social network analysis. *Betweeness centrality* is a measure of how many shortest paths in the network flow through a given node. A node with a high betweeness centrality plays an important role for forwarding information between different clusters of nodes. The *similarity* between two nodes measures the degree of common neighbors. Two nodes have a high similarity if they share a large number of common social acquaintances with respect to their total number of social acquaintances. The intuition behind this is that if the similarity between a potential forwarder F and the target node is high, F may be a good candidate for the next hop because it is very likely that F meets the target node. Both metrics are approximated in a distributed manner, and their weighted sum is utilized for forwarding decisions. Different betweeness measures are also used by Cuevas et al. for *content distribution* in DTNs [CJGS08].

Publish-subscribe systems feature nodes which publish information and subscriber, which receive information according to some criteria. Costa et al. employ the *change degree of connectivity* for routing in a publish-subscribe system, based on DTNs [CMMP08]. The change degree of connectivity is high, if a node has frequent encounters with many neighbors. Such nodes correspond to hubs in a social network.

Hui et al. use distributed community detection mechanisms to infer local communities [HCY08]. The *hierarchical structure of communities* is subsequently used for forwarding decisions. Yoneki et al. use similar mechanisms to create an overlay for a publish-subscribe system based on delay tolerant networks [YHCC07]. They identify nodes with a high centrality (i.e. hubs) within the detected communities and structure the overlay to utilize them. Boldrini et al. use detected communities to *disseminate data* in opportunistic networks [BCP08].

2.3.2 Utilizing Information about Social Context

Miklas et al. have investigated the benefit of incorporating knowledge about social context in routing decisions [MGC+07]. They simulated different simple routing protocols using a globally known social network. These protocols forward a message from a source to the first f encountered friends of either the destination or the source. However, these first hops may forward the message only directly to the destination, which leads to maximum path length of two hops. They report a significant performance benefit compared to protocols which do not use information about social context with respect to the network bandwidth used. Furthermore, the authors show that other applications like firewalls or p2p-systems may benefit from exploiting information about social context.

Recently, Mtibaa et al. have performed a similar study, but implemented more complex multi-hop routing protocols, which exploit knowledge about the social context of the destination node [MCL+08]. The protocols were simulated based on measurements of human mobility, using a self-reported social network, i.e. a network obtained by asking the participants about their social ties. Their protocols focus on selecting a next hop that is closer to the target in the social network in terms of the shortest path. Furthermore, they evaluated protocols which favor nodes with a high centrality. Their experiments show that the combined utilization of both metrics yields results that are close to the optimum while requiring relatively low bandwidth.

Another class of mobile applications for wearable devices detects the (physical) presence of specific devices and informs the user accordingly. For example, *Serendepty* [EP05] uses a database with profile information about the users. If the application detects the presence of another user, it queries the database for profile information. The profiles are compared using a scoring system and if a sufficient match is detected, the software alerts both users about their similarity. The goal is to initiate contacts between users who would otherwise be unaware of their similarity. Pietilainen et al. have generalized the idea of Serendepty to a general architecture for developing applications using a social network and corresponding profiles for initiating contacts between nearby users [POL+08].

Socio-Graph [APV08] is a media browser which focuses on the exploration and sharing of personal media. Social-Graph collects and exploits knowledge about social context to filter media items. For example, a user can filter media dependent on the strength of the social relation with the publisher. The knowledge about social context is collected from human-carried GPS devices.

2.3.3 Conclusion

As we have shown in this section, there are a growing number of applications exploiting social context or the structure of social networks. Thus, respective simulation environments are required that focus specifically on the accurate simulation of sociological aspects. In the following section, we will give some background information on the methods used to perform simulations in this area.

2.4 Mobile Wireless Network Simulation

In this section, we discuss two common ways to simulate mobile wireless networks: Simulation based on recorded measurements of human mobility and the use of mobility models.

2.4.1 Simulation based on Mobility Traces

One way to simulate mobility is by using traces of human mobility recorded in real environments. Essentially, such mobility traces contain the begin- and end-timestamp for each time interval in which two arbitrary nodes were in wireless range of each other. This time interval is denoted as a **contact** between two nodes. Some traces additionally contain GPS coordinates [CBD02]. Different kinds of traces exist that differ in the way they determine if two nodes were in contact at a specific instant of time. The Dartmouth College has started a project which aims to provide an online repository for mobility traces [Dar].

Many publicly available traces [HH06a, CHD+07, EP06] are access point (AP) logs, i.e. recorded time intervals in which nodes were connected to different APs. To yield device-to-device contacts, the following simplification is used: Two nodes are considered to be in contact at time t if both nodes are connected to the same AP at t. Clearly, this is only an approximation of real mobility. Despite being in range of the same AP, both nodes may still be out of range of each other. Furthermore, two nodes may be in range of each other but out of range of any AP. In this case, the contact will not be recorded at all. Traces created from AP logs are called **AP-based traces**.

On the other hand, **direct contact traces** [CHD+07] measure human mobility in a more realistic way. Wireless devices running a special mobile application are distributed among participants of an experiment. The devices perform scans at periodic intervals to discover and record the presence of other nodes. For example, Hui et al. distributed Intel iMotes among 41 participants of the INFOCOM 2005 conference [HCS+05]. Each iMote performed a five second long Bluetooth based 'inquiry' scan every 120 seconds. Other nearby devices responded with their MAC addresses to such an inquiry. Each newly encountered node is placed in an 'in-contact' list and the begin of the contact interval is recorded. If a node in the 'in-contact' list misses to respond to a subsequent inquiry, the connection is considered to be lost and a timestamp is recorded, which approximates the end of the contact.

Still, even direct contact traces are only an approximation of human mobility. Devices miss to record very short encounters with high probability due to the relatively long scan period. However, a shorter scan period would drain the batteries of mobile devices too fast. This is the reason why similar experiments used even higher scan intervals, e.g. 300 seconds in experiments performed at the MIT [EP06].

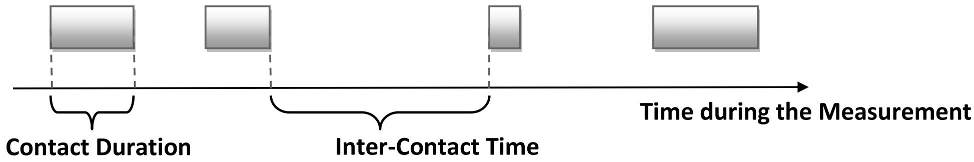


Figure 2.4: Contact duration and inter-contact time for a pair of nodes (cf. [CHC+05]).

Both, the contact duration and **inter-contact times** are used to characterize different mobility scenarios [CHC+05]. An inter-contact time is defined as the time elapsed between two successive contacts of two fixed nodes. Figure 2.4 shows the relation between contact duration and inter-contact time. The distribution of contact durations represents the *capacity* of an opportunistic network. Because in sparsely connected networks long delays between contacts are most common, the *aggregated* inter-contact time distribution (short: inter-contact distribution) is of special importance. The inter-contact distribution characterizes the frequency of contact between two arbitrary nodes. Hence, it represents the *delay* of an opportunistic network.

For a long time, the inter-contact distribution has been assumed to be short-tailed (i.e. exponential) [Cai07], which means that nodes meet frequently and long time intervals without any contact are very rare. However, Chaintreau et al. have recently investigated the CCDF (complementary cumulative distribution function) of inter-contact times using eight different traces, both AP- and direct contact based [CHD+07]. They report an interesting observation: The tail of all traces is well approximated by a power-law until an order of magnitude of about one day. Power-law coefficients are typically found to be in the range $\lambda \in [0.2, 0.6]$. The discovery of a heavy-tailed inter-contact distribution has the consequence that 'simple' forwarding algorithms that do not keep a history of previous contacts with other nodes do *not* have a finite expected delay. Intuitively, this is due to the relatively high probability of long periods without connectivity between nodes compared to a short-tailed (e.g. exponential) distribution.

Karagannis et al. later confirmed and refined these results [KBV07]. They discovered that the CCDF of inter-contact times is characterized by a power-law until a characteristic time of about half a day, followed by an exponential decay. This *dichotomy* seems to be an invariant of human mobility, independent of the length or source of the trace. However, the length of the exponential tail seems to be approximately proportional to the total time a given trace was recorded [CHD+07].

Figure 2.5 shows the described dichotomy for the above described trace from the INFOCOM 2005. The probability that an arbitrary inter-contact time is longer than t (y-axis) over a range of times t (x-axis) is depicted on a log-log scale. A distinct power-law until about half

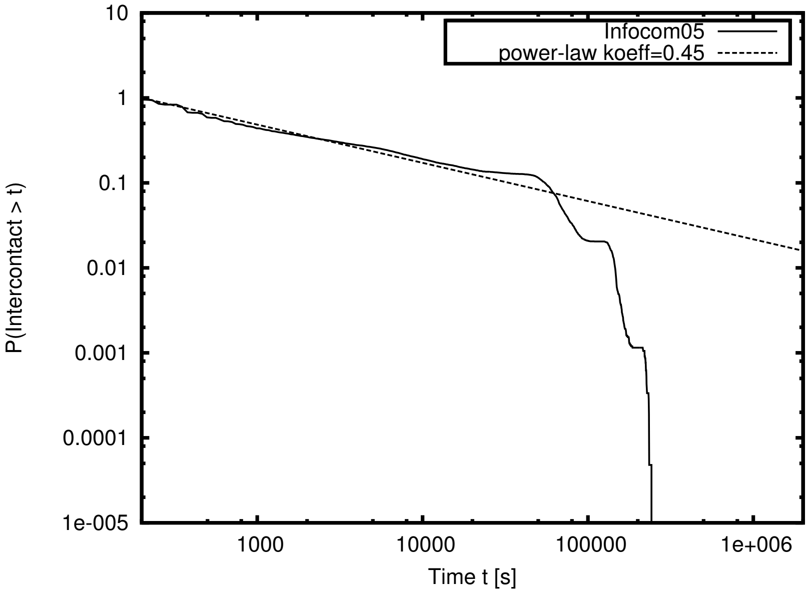


Figure 2.5: Inter-contact CCDF of a trace, measured during the INFOCOM 2005 conference.

a day is observable, followed by an exponential degradation. Because the experiment lasted only for three days, the exponential degradation is very short compared to other traces.

Although real traces of human mobility can be used for the evaluation of mobile applications, they have their limits. Especially direct contact traces are, unfortunately, very scarce and all publicly available traces feature only a small number of nodes. Furthermore, traces are typically of short duration and measure only specific scenarios such as conference or campus environments. No variation of parameters is possible, like the number of nodes or distribution of speed [MM08]. Thus, we need appropriate and realistic mobility models which compensate for these disadvantages.

2.4.2 Simulation based on Mobility Models

Essentially, a mobility model consists of a set of rules which define the movement of mobile nodes on the simulated area. From this information, a *connectivity graph* [MCL$^+$08] emerges. A connectivity graph is a temporal network [KKK00], i.e. edges between nodes change over time. An edge between two nodes in the connectivity graph exists at time t if both nodes are in contact (i.e. in wireless range) at time t. In most situations, only the connectivity graph is required for simulation. However, applications exploiting geographical information (e.g. geographical routing) require the actual geographical position of each node.

Recently emerging *connectivity models* [KPR06] try to simulate a connectivity graph directly. However, such models are very rare because it is difficult to make any assumption on realism at this level. Most connectivity models are based on results from real connectivity traces [MM08].

Synthetic Mobility Models

The most simple and widely used mobility models are *random mobility models* [CBD02]. One prominent model of this class is the Random Walk mobility model, first described by Einstein as the Brownian motion [Ein56]. In this model, nodes move at a random speed and towards a random direction for a certain distance or time. After that, a new random speed and direction is chosen for the next trajectory. The Random Waypoint Model extends the Random Walk mobility model by adding **pause times**, i.e. after each trajectory, nodes stay for a certain duration at a random point on the simulation area before choosing a new speed and direction [JM96]. Numerous further variants of random mobility models have been proposed [CBD02].

Although being simple, random mobility models are not a very realistic approximation of human mobility. After the discovery of the heavy tailed inter-contact distribution of human mobility (see Section 2.4.1) many of the early random mobility models were invalidated by this result because they are characterized by an exponential inter-contact distribution. Because of this surprising invariant, most recent mobility models are validated by comparing their inter-contact distribution to the characteristic of real mobility traces.

To reflect human mobility in a more realistic way, numerous improved mobility models were proposed. The Gauss-Markov model [LH99, CBD02] concentrates on avoiding sudden stops or sharp changes in direction. Speed and pause times may be selected to reflect real scenarios [KKK06]. For example, cars have different mobility patterns compared to pedestrians [DNS08, HFB08]. Other models constrain the movement of nodes, e.g. by incorporating streets like in the city section model [Dav00] and the manhattan/freeway mobility model [BSH03]. To improve the realism further, streets can be extracted from real maps [ATB08]. In addition, humans do not walk through walls, thus other models introduce obstacles to constrain the movement of nodes [JBRAS03]. Obstacles not only have an impact on the movement of humans, but also on radio propagation. Hence, advanced radio propagation models have been proposed [SHR05, SR06]. Because such models concern themselves with geographical aspects, we call them **geographical mobility models** [Her03]. Geographical mobility models are typically very scenario-dependent and concentrate on modeling 'micro-mobility'.

So far, all discussed mobility models have in common that the movement of nodes is completely independent of each other. Since the movement of humans is strongly influenced by their social relationships, the new generation of **social mobility models** emerged recently. We will discuss the research in this area in Chapter 4.

Group mobility models concern themselves not with the movement of individual nodes, but with the movement of groups of nodes. For example, in the column mobility model, groups are represented by nodes which move together in a line [CBD02]. Nodes in the reference point group mobility model [HGPC99] move relative to a reference point, which represents the logical center of the group. The reference point moves around the simulation area while the nodes within the group move in a random fashion around the reference point. In the boids model [Rey87], nodes move as a flock based on three simple rules. Many additional group mobility models exist [MM08, CBD02]. Because they focus on node movement on a microscopic scale, we consider group mobility models as belonging to the class of geographical mobility models.

Trace-based Mobility Models

Trace-based mobility models concentrate on the enrichment of a synthetic mobility model with data extracted from real mobility traces.

The WLAN mobility model [TG05] uses data retrieved from an AP-based trace recorded at the ETH in Zurich. The authors divide the simulation area into cells. Each cell corresponds to an area captured by an AP. Nodes are initially placed using a uniform random distribution. Subsequently, nodes move probabilistically between cells. Only cells are visited that were also visited in the trace. The probability to move to a neighboring cell is calculated based on data extracted from the trace. Nodes stay at cells for a certain time drawn from the distribution of the session duration of the corresponding AP.

Kim et al. propose a very similar approach by using an AP-based trace from the Dartmouth College [KKK06]. In addition, a speed distribution is extracted from the trace. For each node trajectory, a random speed is chosen from this speed distribution.

Other models use data retrieved from user surveys to create probabilistic mobility traces [ZHL06, BRS+04, HMS+05]. Essentially, they are similar to the examples discussed above.

Conclusions

As we have discussed, a multitude of different mobility models exist that focus on different aspects. Researchers have to make a decision on which model should be used for his individual requirements. If these requirements change or if a single model does not satisfy them, this creates a problem. Thus, in order to avoid this situation and to obtain a simulation which is as realistic as possible, it is important to design more general models that can be parameterized for specific target scenarios.

Chapter 3

Requirements Analysis for Realistic Mobility Models

One of the primary goals of this thesis is to design a mobility model which is realistic. In this chapter, we elaborate on the implications this goal has on the requirements of the model.

As shown by many authors, the mobility model has a strong impact on the simulation results [PHO02, JBRAS03, CBD02, BSH03]. Thus, it is sensible and important to develop mobility models which are as realistic as possible. However, the level of detail must be limited to some point. Too much details increases the simulation run-time, complexity of the implementation and hence the probability of bugs [HBE+01].

Thus, we identify a small set of properties of human behavior which we consider important enough to be captured by a realistic social mobility model. In this process, we focus on properties related to social context and avoid to constrain properties related to geographical mobility models, like speed or pause times. Note that we already have discussed that real mobility traces are characterized by an inter-contact distribution which features a power-law, followed by an exponential degradation. Besides this empirically measured characteristic, human mobility is shaped by the following mobility patterns.

3.1 Social Context

An important property of human behavior is that we are strongly influenced by our social context. In this section, we advocate the use of a social network as input and its accurate reflection, as discussed in Section 1.2. We have identified the following reasons:

- **Social networks directly influence human mobility.** Only few researchers have investigated the influence of the underlying social network on the connectivity graph. Mtibaa et al. have compared a mobility trace of a conference environment to the self-reported social network of the participants [MCL+08]. The degree of nodes in the social network correlates to the average degree of nodes in the connectivity graph. However, the average node degree in the connectivity graph is significantly smaller compared to the node degree in the social network. Strong correlations have also been observed in terms of node centrality. These results show that important structural properties of the social network are mapped to the connectivity graph.

Mtibaa et al. report further that contact durations between social acquaintances are typically a magnitude longer compared to contact durations between nodes which did not report themselves as friends. Hui extracted social networks from multiple mobility traces [Hui07]. He inferred social relations by using a formula based on the frequency and duration of contacts. He used a community detection algorithm to compare the communities of the extracted social network to the actual communities and observed a good match. Furthermore, he reports that the contact duration correlates with the contact frequency, that is, nodes which meet often typically also spend much time in contact.

The work of these authors suggests that the social network has a strong impact on the mobility patterns of humans.

- **Model independent simulation.** The fact that so many different network models exist (cf. Section 2.2.3), shows that having a model-independent simulator is a key feature. Using an arbitrary social network as input, possibly obtained from a connectivity graph, and having a simulator that reflects this network makes the simulator independent of any specific properties of human social behavior. It can reproduce *any* properties found in the input network. Thus, it enables the integration of future findings in this area without changing the simulator itself. In additions, this approach provides much more control over the simulation.
- **Reference for comparison.** As discussed in Section 2.3, existing work shows that mobile applications can benefit from exploiting knowledge about the social context of users. Much of the results assume a globally known social network. However, we argue that if no knowledge about the underlying social network is known, mobile applications still may detect and collect (partial) information about the social context of nodes. For example, individual nodes may maintain statistical information about the contact frequency, duration, and/or inter-contact times with other nodes and hence infer knowledge about their social context. This information may be shared with other nodes in an opportunistic manner, thus acquiring partial knowledge about the social network. To evaluate algorithms and protocols which collect and/or exploit knowledge about the social context of nodes, researchers should be able to compare their results to the underlying social network. The reflection of an input social network provides such a reference.
- **Integration of realistic structural properties.** As discussed in Section 2.3, mobile applications may exploit structural properties of the underlying social network. For

example, we have seen that some applications utilize nodes with a high betweenness measure, which is a characteristic shown by nodes connecting different social communities. However, it was discovered already over three decades ago that such 'weak' links between social communities play a vital role in dissimating information in social networks [Gra73]. Recent work confirms that weak links in the social network are also of high importance on the level of connectivity for content dissemination systems [IC09].

To enable the research of mobile applications exploiting the structures of social networks, a social mobility model must create mobility based on these structural properties. An input social network allows the integration of *arbitrary* structural properties into the mobility scheme.

3.2 Active Social Relationships

Intuition tells us that relationships between humans are not equally important at each point in time. For example, two individuals who share a 'colleague' relationship meet typically during work hours. Two individuals sharing a 'family' relationship, on the other hand, rather meet in their free time, but not during work hours.

The analysis of real mobility traces confirms this intuition behind human social behavior: Very heterogeneous inter-contact times have been observed between the same pair of nodes [CLF07]. In other words, two individuals sometimes meet very frequently and sometimes meet not at all during certain periods of time. This motivates the following two definitions:

Definition 3 (Active Social Relation/Acquaintance) *We call a social relation* ***active*** *at time t, if two individuals typically meet at t with a significant probability. A social acquaintance u of a node v is called active social acquaintance at time t, if the social relation between u and v is active at t.*

Note that since it is difficult to analyze human social behavior analytically, it is not possible to quantify what 'significant probability' means in the general case. Nevertheless, based on this and the following definition, we advocate that the probability for connectivity between two nodes is not uniformly distributed over time.

Definition 4 (Active Social Context) *The set of all active social acquaintances of a node v at time t is called the* ***active social context*** *at t.*

The integration of this concept into a mobility model is of significance for mobile applications which exploit knowledge about the social context of users. For example, we consider an application which collects statistical information about the average inter-contact time for all encountered nodes. Assume that a message m must be transmitted from node A to node B. A knows that the average inter-contact time between A and B so far is 1 hour. Yet, this does not necessarily mean that A and B meet roughly once every hour at any given time. Instead, if A and B are office-mates, they may meet on average every 20 minutes *during* office hours, but only very seldom outside the office, which may result in about 1 hour of *average* inter-contact time. Based on this relatively short average inter-contact time, the application running on node A decides to directly transmit m from A to B i.e. to wait for the next contact opportunity with B. However, if the time of this decision is Friday night, the time until the next contact may be more than two days with a high probability since A and B do not work at weekends. Thus, it might be better to utilize another node as the next hop for m. Instead of only measuring the *average* inter-contact time, our example application could measure the average inter-contact time for different periods of time. This would allow the application to infer that the inter-contact time between A and B is relatively high between Friday evening and Monday morning. More generally, an intelligent mobile application may keep track of the active social context over time.

3.3 Spatial Regularity

There seems to exist a strong dependency between individuals and specific locations. In particular, the probability to visit an arbitrary location is not uniformly distributed as many mobility models assume. Instead, each individual is associated with several locations of significance which he visits frequently and regularly, like the working place, home, etc. Thus, humans are characterized by a spatial regularity. This motivates the following definition based on Adams et al. [APV08]:

Definition 5 (Social Sphere) *A social sphere of a node defines a set of important locations for the corresponding user which he visits frequently and regularly.*

Locations of the social sphere also act as a place where individuals meet other individuals with whom they share social relations and thus play a geo-social role.

Some research focuses on the study of human spatial regularity. Gonzales et al. have investigated the reappearance of humans at specific locations in the trajectories of 100000 mobile phone users [GHB08]. For each location, they ranked the number of visits where L represents the L-most visited location. They report that the probability to find a user at a location of rank L is well approximated by $P(L) \sim L^{-1}$ independent of the total number of locations a user has visited. This *heavy-tailed distribution* suggests that a typical human individual spends most of his time at a few locations while visiting a large range of locations with low probability.

Hsu and Helmy [HH06b] have performed a similar investigation using several different mobility traces. They have ranked the time users spend associated to different APs and calculated the average over all users. The results are shown in Figure 3.1. In particular, an average individual spent more than 95% of his online time associated with only five APs. Very similar results were observed in a study conducted at the ETH Zurich [TG05].

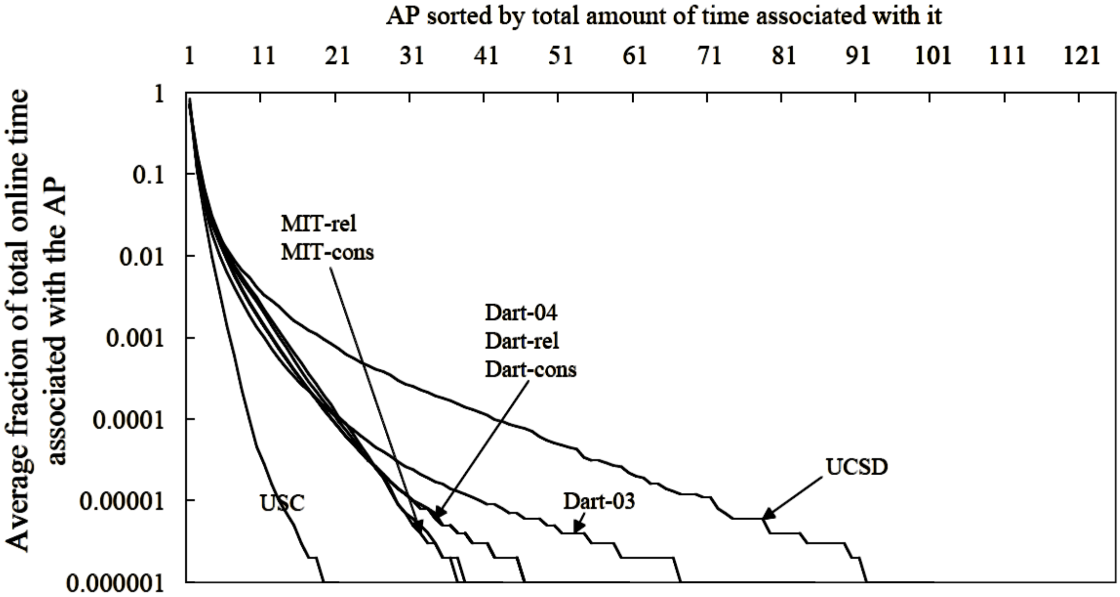


Figure 3.1: Average fraction of the time a user spent associated with a given AP for different mobility traces (semi-log scale) [HH06b].

These results confirm the existence of social spheres. Other work aims at actually collecting the social sphere of users by analyzing GPS-based mobility traces [GBNQ06, CHK05].

If a node spends most of its time at a few locations, it will often meet the same mobile nodes, which also tend to follow this pattern. Such regularly emerging locality-induced social groups may be exploited by mobile applications [JS09].

3.4 Temporal Regularity

Humans are also characterized by a temporal regularity. The probability that a user is located at a specific location is not constant over time but very heterogeneous. In addition, this temporal dependency seems to be periodic, which leads to a *temporal regularity*. More precisely, there exist time gaps t_g such that the probability that an individual reappears at a certain location after t_g is significantly increased compared to other time gaps. This confirms the intuition that humans follow daily and weekly schedules.

To study the temporal regularity in mobility traces, Hsu and Helmy [HH06b, HH05] have defined a metric called *network similarity index* (NSI). To study the tendency of users to reappear after a certain time gap (e.g. 24 hours), the authors evaluated if a user was located at the same location for two points in time t and $t + t_g$. The authors evaluated this for each minute during the measurement. The *location similarity index* for a user and time gap is then defined as the fraction of times the statements holds. Finally, the *network similarity index* for a specific time gap is calculated as the average location similarity index over all users for this time gap. Figure 3.2 shows the NSI (y-axis) over a range of time gaps (x-axis) for different mobility traces. Distinct peaks after multiples of a day are observable. This result suggests the tendency of users to return to certain locations after a period of one day with high probability. Other work shows similar results [GHB08, SA03].

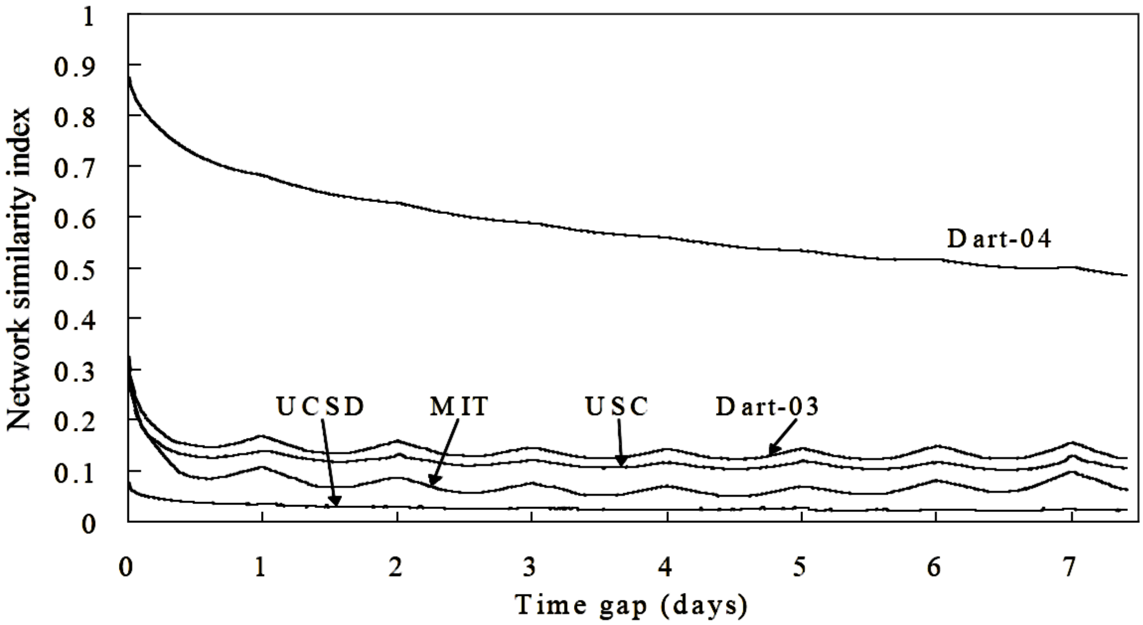


Figure 3.2: Network similarity index for a range of time gaps using five different mobility traces. The peaks suggest a period of one day [HH06b].

If the movement of mobile nodes is based on repetitive patterns, the connectivity of two nodes can be predicted with a certain probability. Thus, mobile applications may keep track of the connectivity over time and identify time intervals in which there is a high probability to encounter specific nodes. This knowledge may then be exploited, e.g. for forwarding decisions.

3.5 Group Movement

Frequently, humans do not move independent of each other from location to location, but rather move as a group. We argue that group movement is a realistic mobility pattern and may be exploited by mobile applications, e.g. by identifying groups and utilizing their combined resources.

Several examples of mobile applications which explicitly consider group movement exists. For example, Thomas et al. [TGK06] propose a routing protocol in a mobile network in which most nodes move in groups. These groups are identified by a distributed algorithm and utilized for routing on the group-level. The authors report improved performance compared to traditional approaches. As another example, the Social Rope [NYBK06] runs on wireless mobile phones and keeps track of groups in which users move. If a group member gets lost, the Social Rope informs the other users. This resembles a rope, which ties together a group of people and gives immediate physical feedback if a group member moves away.

To enable the research of mobile applications which consider moving groups of individuals, a realistic mobility model should feature group mobility. To capture different mobility scenarios, the number of group movements should be controlled by a parameter.

3.6 Conclusions

In this chapter, we have identified several important *realistic* human mobility patterns that lead to the following requirements to achieve our goal of realism:

- Social context must be explicitly simulated. The global social context of all nodes is provided by using a social network as input for the model.
- The probability for connectivity between two nodes must not be constant over time since in the real world, people meet during certain periods of time with increased probability.
- The movement of mobile nodes must be characterized by spatial and temporal dependencies.
- Nodes must not always move completely independent of each other but rather move as a group. A parameter should control the number of group movements.
- A realistic mobility model should satisfy the characteristic dichotomy of the inter-contact distribution measured in real mobility traces (cf. Section 2.4.1).

A mobility model that satisfies these requirements captures a strong degree of realism. We have also discussed examples of applications which may utilize these patterns. This shows their importance in evaluating such applications. The identified requirements are later implemented in the construction of the model (Chapter 5). In Chapter 7, we show that the model satisfies these requirements.

Chapter 4

Related Work

In recent years, several mobility models incorporating the fact that social relations influence the movement of humans have been proposed. In this section, we discuss these models. However, several of the models focus rather on the simulation of (unstructured) communities than individual social relations. In each of the following sections, we first illustrate the main concepts of the different approaches. Then, we discuss their realism as well as advantages and disadvantages. In addition, we discuss the fulfillment of the requirements identified in the last section. To conclude this chapter, we discuss the implications of the existing social mobility models on the design of our proposed model.

4.1 First Social Mobility Model

Herrmann proposed the first mobility model which focuses on the simulation of sociological aspects [Her03]. The model takes a non-weighted network as input and aims to reflect this network in its mobility scheme by producing frequent encounters between nodes which share a social relation.

The simulation is conducted in multiple simulation periods which are time intervals of constant length. Each clique of nodes in the social network is associated with a time interval in which its members meet at a fixed location within the simulation area. Based on this schedule, nodes move from location to location to meet all cliques of whom they are member over the course of one simulation period. Subsequently, the predefined schedule is repeated.

The author validates his model by the creation of an output social network, which is extracted from the connectivity graph based on the number of encounters. By comparing input and output social network, the author observes that the structure of the social network is preserved quite well.

The model set the first steps towards the integration of sociological aspects into a mobility model. As will become apparent in the following sections, it successfully triggered many further models. However, the model is very simplistic. We argue that a deterministic model based on fixed schedules is not very realistic. In particular, the model is not suited to evaluate mobile applications which exploit spatial or temporal regularities. Because each node always meets again after the same constant time interval, the inter-contact distribution cannot match characteristics found in real mobility traces.

4.2 Community-based Mobility Model

The community based mobility model (CMM) employs a social network generated using the caveman model (cf. Section 2.2.4) to create its mobility scheme [MHM04, MM06, MM07]. Each *pair of nodes* in the social network is assigned a constant weight, which represents the strength of the social relation (cf. Definition 1). The weight is zero if two nodes do not share a social relation.

An algorithm proposed by Newman and Girvan [NG04] is used to identify communities in the social network and to partition the set of nodes into non-overlapping communities. Each community is placed initially on a random cell in a grid, which represents the simulation area. At each simulation tick, every node is *associated* with a single point on the grid. This point is called the *goal* of the node. Nodes always move directly towards the direction of their associated goal with a constant velocity until they reach it. Subsequently, a new goal (possibly within the same cell) is selected based on the *social attraction* of each cell. The social attraction a cell C exerts towards a node v is calculated as the average weight between v and nodes currently associated with C. If no node is associated with C, the social attraction of C is zero per definition. After the calculation of the social attraction for each cell, a random point within the cell that exerts the *strongest* social attraction is chosen as the next goal.

To implement the concept of active social relationships (see Section 3.2), the social network is periodically regenerated (e.g. every 8 hours). After such a social network reconfiguration, all nodes in a newly formed community are immediately associated with a new goal within the same random cell on the grid.

The authors have validated their model by showing that its inter-contact distribution matches the characteristic of real traces, but depicted their results only for roughly one third of a day. This does not show if it matches the dichotomy observed in real traces. Furthermore, the CMM does not feature temporal regularities. The periodic reconfiguration of the social network leads to a sudden change of the active social context of all nodes at the same time. We consider both this and the fact that the different generated active social contexts are completely unrelated to be not very realistic.

The model does not focus on the reflection of the social network. However, we have performed experiments and observed that the social network is reflected rather poorly. We have identified the following reasons:

- The cell selection algorithm creates the following problem: If the last node leaves a cell C, no other node will move to C again because the social attraction towards C is zero for all nodes. We have performed many simulations with different parameters and observed that this leads to an monotonic decrease of the number of groups until only two or three groups are left, independent of the original number of communities. This leads to a 'blurring' of the social network structure between and within communities on the level of connectivity because they are deemed to be grouped together until the next reconfiguration of the social network.
- In other cases, some or all nodes did not move at all because their currently associated cell is the cell with the highest social attraction. This leads to a (possibly infinite) stationary behavior, which prevents further meetings between many social acquaintances.
- Social relations between members of different communities are not properly reflected. Consider the case that a node A has only a single social acquaintance B in an external community (cave) C_e. Because most of the time B is co-located with members of C_e, B's cell typically exerts a lower attraction towards A than other cells to whose members of A's own community are associated. Thus, A meets only very seldom (or even never) with B despite the fact that their relationship may be as strong as any relationship within A's community.

4.3 Home-cell Community-based Mobility Model

Because the movement of humans is not only driven by social relations but also by physical locations, Boldrini et al. proposed the home-cell community-based mobility model (HCMM) [BCP07], which is based on the CMM. In this model, all members of a community share a single *home cell*, which corresponds to the cell where the community is randomly placed after each social network reconfiguration.

In the HCMM, goals are selected based on a probabilistic goal selection algorithm. The probability that a node v chooses the cell C as the next destination anchor is proportional to the average weight between v and nodes *having their home at C*. Note that this is in contrast to the CMM, where the attraction depends on the current node association. If a node reaches an external cell, i.e. the home cell of another node, it stays there until the next time step with probability p_e and returns to its home cell with probability $1 - p_e$. Thus, p_e controls the average pause time $p_e/(1 - p_e)$ at an external cell. The authors show that their model creates a heavy tailed inter-contact distribution. However, this distribution does not resemble the dichotomy observed in real traces.

Like the CMM, the HCMM does not focus on the reflection of the social network. Because the members of a community share a single home cell, the structure within a community (e.g. hierarchical) is completely 'blurred' because all nodes of a community share the same attraction towards their home cell and therefore meet each other with the same probability.

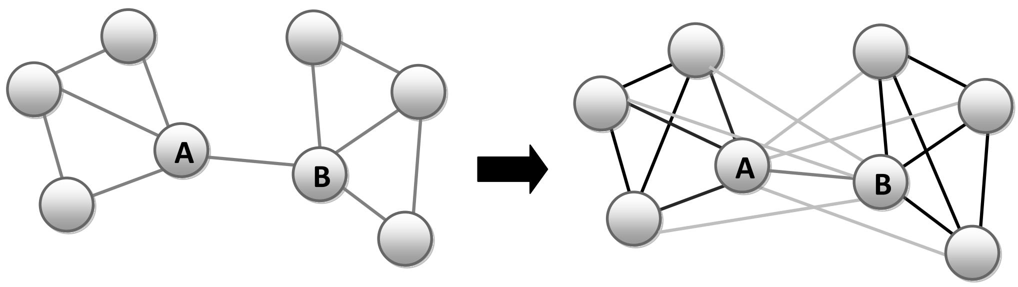


Figure 4.1: Reflection of the social network in the HCMM.

We use the excerpt of a social network, shown in Figure 4.1 (left), to illustrate further reasons why the model does not reflect the social network properly. We assume equal weights for all social relations. Based on the shown social network, node A will stay at the home-cell of B's community with a probability which is significantly less than the probability to stay at the home-cell of A's community. This is due to the fact that A has no social relation with most of the members of B's community, which results in a weak home-cell attraction. Hence, A meets B very seldom compared to other nodes of its community, independent of the strength of their relationship. Furthermore, A will meet each member of B's community with equal probability at B's home cell, despite the fact that A has no social relation with them. Similarly, A will meet members of its own community with the same probability at their common home cell, independent of their relationship. Figure 4.1 (right) shows schematically the effect of this behavior on the connectivity graph. The brighter we have depicted the social relation, the lower the contact probability. As shown, the structure of the social network between nodes in the same community is 'blurred'. The contact probability between A and B is too weak with respect to the strength of their social relation, and A meets with equal probability with each of B's community members and vice versa.

4.4 Time-variant Community Mobility Model

The time-variant community mobility model (TVC) [Hsu08, HSPH08] and its predecessor [HSPH07] aim to reproduce the spatial and temporal regularities observed in real traces of human movement. In the TVC model, the time is structured into *time-periods*. Different time-periods may be defined and chained together in a recurrent structure, e.g. TP1, TP2, TP3, TP1, TP2, TP3, TP1, . . . In each time period, a number of communities are defined. A community is a square-shaped geographical area on the simulation plane. Communities are very similar to the concept of home-cells in the HCMM. Node movement is created by a sequence of *epochs*. An epoch involves the random movement of a node within the geographical boundaries of a community, similar to the random waypoint model. The probability to perform an epoch in a specific community is dependent on the node and the current time-period. The number and geographical assignment of communities may differ in different time-periods.

The use of this mobility model involves extensive configuration. For each time period, the user of the TVC model must specify the following (among others):

- The duration of the time epoch
- The number of communities and corresponding geographical assignment
- For each node v and each community c:
  - The probability that v performs an epoch within the geographical boundaries of c
  - Average pause time for epochs in c
  - Average epoch length for epochs in c

While time consuming, this configuration allows the TVC model to capture the spatial and temporal regularities as measured in real traces quite well. The authors have also shown that their model captures the dichotomy of real mobility traces for several configuration scenarios.

However, the use of discrete time periods to create spatial and temporal regularities leads to a sudden change of spatial and temporal behavior for all nodes at the same time. In the real world, however, this change happens more gradually. Thus, we employed a different mechanism in our model which leads to a continuous change of spatial and temporal behavior.

4.5 Two-level Social Mobility Model

In the two-level social mobility model [GGP08], nodes are partitioned into disjoint communities. They move between fixed locations on the simulation area, similar to the CMM. However, the probability for a node to choose a certain location as the next goal is based on a fixed heavy-tailed Zipf distribution. Nodes of the same community share the same probability distribution, which leads to frequent meetings between nodes of the same community.

Each location is partitioned into several so called *aggregation points*. During their stay at a specific location, nodes move between aggregation points. Pause times on the location and aggregation point level are selected based on a log-normal distribution, with a smaller mean value for the latter. The model introduces a circadian rhythm by stopping all nodes for a certain period of time (i.e. at night).

By only measuring contacts created at aggregation points (i.e. neglecting random encounters on the way), the model captures the dichotomy of the inter-contact distribution of real mobility traces quite well. The idea of a heavy tailed location visiting distribution should resemble the spatial regularity of human mobility quite well, though the authors did not evaluate this property explicitly. However, the model does not capture temporal regularities, nor does it focus on the reflection of the communities.

4.6 Sociological Interaction Mobility for Population Simulation

This model [BLdAF06], which we call SIMPS in the following, is based on two sociological findings:

- Intrinsicality: Humans have a fixed sociability need dependent on intrinsic factors like social class and age.
- Interactivity: Humans try to satisfy their level of sociability by encountering social acquaintances or forming new social acquaintances.

SIMPS is based on a weighted social network generated by the Barabasi-Albert model (see Section 2.2.4). Fixed sociability needs of nodes are generated based on a normal distribution. The current level of sociability of a node is calculated as a running average over the number of social acquaintances within a constant radius around the node.

A node always tries to satisfy its sociability need by alternating between two states: *Socialize* and *isolate*. In the socialize state, a node is attracted by social acquaintances. In the isolate state, a node is repulsed by strangers. If the current level of sociability drops below a certain threshold, a node changes into the socialize state. Similar, if the current level of sociability exceeds another threshold, a node enters the isolate state.

The attraction (repulsion) a node exerts towards another node translates into an attraction (repulsion) vector in the direction of the node. The length of this vector is proportional to the weight (1-weight) of the social relation and inversely proportional to the Euclidian distance. The linear combination of either all attraction or repulsion vectors, dependent on the current state, yields the *social motion influence* of a node. Nodes always move in the direction of the social motion influence in socialize mode and in the reverse direction in isolate mode. In either case, the acceleration of a node is proportional to the length of the social motion influence.

The authors report realistic characteristics of the inter-contact distribution. However, we question the realism of the concept behind nodes which try to constantly keep a fixed number of social acquaintances around them. Other work shows that the connectivity of nodes in mobility scenarios is very heterogeneous [MGC+07]. Furthermore, nodes are in constant motion despite the fact that much stationary behavior has been observed in real mobility traces [HH06b]. Thus, the model does not capture the spatial and temporal regularities measured in real traces. This and the fact that aspects of the geographical model like acceleration (and thus speed) are dependent on factors of the social mobility model, may lead to a challenging integration of geographical mobility models. Although using a social network as input, the authors did not focus on its reflection.

4.7 Discussion

For reasons discussed in Section 3.1, we consider the reflection of a social network as being an important requirement for a social mobility model. Only one mobility model [Her03] satisfies this requirement. However, this model is very simple and thus leaves room for further improvements. No *social* mobility model tries to capture the movement of nodes in groups.

Some models try to focus on producing meetings between nodes based on the shared membership in a community. However, this is a simplification of the structure of real social networks. Hence, we do not consider such an approach as being suited for the evaluation of mobile applications. In addition, no community-based model even tries to show the reflection of the community membership on the level of connectivity.

We adopt the idea behind the reflection of an input social network from the first mobility model [Her03]. However, rather than creating fixed schedules, we focused on a probabilistic approach. To reach this goal, we borrowed the idea of attractions towards nodes which are associated with locations on the simulation area. We did not use an approach based on the TVC or two-level social mobility model because both create meetings between nodes based on the user-level, i.e. shared community membership. Although we think that the idea behind SIMPS is quite interesting, we think that the integration of specific locations to create spatial and temporal dependencies is necessary. However, we consider this integration into a mobility model which depends on the constant, dynamic movement of nodes to create contacts as being very difficult.

We have designed our model to generalize other probabilistic social mobility models [GGP08, Hsu08, MM06] which feature attraction between nodes and/or attraction towards certain locations. This enables the evaluation of different concepts such as user prediction, group movements, social aspects, etc. without changing the used model.

Chapter 5

The Social Mobility Model

In this chapter, we introduce our proposed social mobility model. First, we start by providing an overview of the important model elements and their relations (Section 5.1). We go on by giving a formal definition (Section 5.2). Afterwards, we explain how group mobility is created (Section 5.3) and discuss the factors that dictate the movement of simulated nodes in more detail (Section 5.4). In Section 5.5, we introduce concepts to improve the reflection of arbitrary social networks despite some 'disruptive' factors (e.g. geographical aspects). These concepts can be regarded as an extension of our basic model. Then, we sketch how geographical concepts can be integrated (Section 5.6). In Section 5.7, we discuss several social mobility models which are generalized by our approach and show how our model can be specialized to recreate their behavior. Finally, we summarize the contents of this chapter (Section 5.8).

5.1 Overview

In the following, we take an abstract look at the elements of our model and illustrate the basic idea behind its design. We will discuss the individual concepts in more detail in the subsequent sections.

5.1.1 The Big Picture

We use a social network as the primary input for the model. This network represents the individuals and their relationships whose movements we simulate. For each node in the input social network, a **mobile node** (or short: node) is created assuming that the corresponding user carries a single mobile device. The simulation occurs on a two dimensional rectangular area. Social interactions take place at so-called **anchors** which are placed randomly on the simulation area. Anchors are *abstract locations* that may represent, for example, a bus stop or a coffee table. Nodes move across the simulation area between them. If a node reaches an anchor, it stays there for a specific duration. This creates connectivity between nodes co-located at the same anchor, which we call a **meeting** between those nodes.

After a node stayed at an anchor for a specific duration, it starts to move towards a new destination anchor. Probabilistically, the next destination anchor is not selected by a single node independently, but by several nodes at the same time. Subsequently, those nodes move together to their new destination, which creates group mobility. The movement of nodes is dictated by different types of *attractions*. Abstract locations, represented by an anchor, exert an attraction in order to create spatial and temporal regularities (cf. Chapter 3). More precisely, nodes prefer to visit a small subset of all anchors and they reappear after periodic time intervals with high probability. Nodes are also attracted by other nodes to yield frequent meetings between social acquaintances. In some cases, it is necessary that nodes are repulsed by each other in order to reflect the input social network appropriately (cf. Section 1.2).

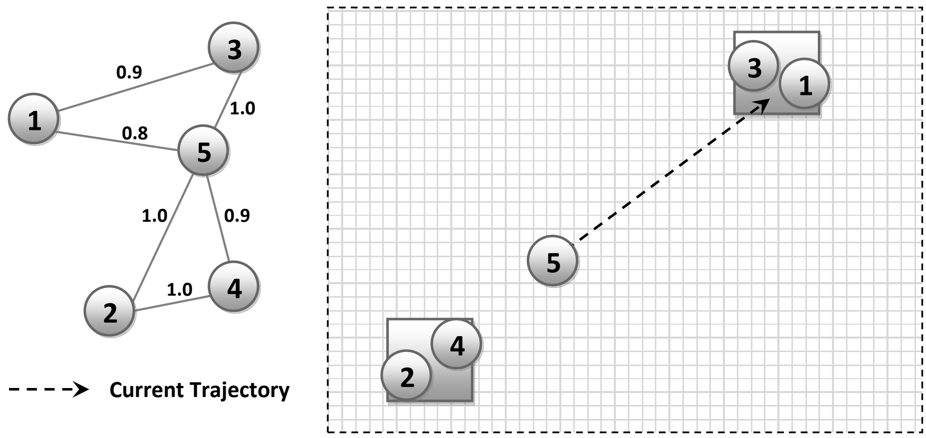


Figure 5.1: Example for a typical simulation scenario with nodes (circles) and anchors (squares).

An example for a typical snapshot during the simulation of the model is shown in Figure 5.1. The input social network is shown at the left side. The simulation area with anchors (squares) and nodes (circles) is shown to the right. Nodes 2 and 4 and nodes 1 and 3 are currently meeting at an anchor. This represents a typical situation because both pairs of nodes share a strong social relation. Node 5, however, is moving from the lower to the upper anchor. This is a typical situation as well since node 5 has strong ties with both pairs of nodes.

5.1.2 Idea behind the Model

The mobility scheme of our model is based on the following real-world abstraction. As introduced in Section 5.1.1, there are essentially two types of attractions that dictate the movement of humans. First, humans are attracted by certain locations such as their home, working place, or favorite restaurant. The strength of this attraction is not constant over time but very heterogeneous and periodic. For example, assume that a team of colleagues has a weekly team meeting. Every week during the time of the meeting, the participants are co-located at the meeting room with a high probability. One possible interpretation of this behavior is that such locations exert a weekly attraction peak towards some specific individuals, in this case the participants of the meeting. However, during other time intervals (e.g. at night), locations do not exert any significant attraction. The attraction peak follows a weekly pattern, hence, it is periodic. This heterogeneous, periodic attraction can be regarded as the reason for the temporal regularity of human movement (cf. Section 3.4). Second, humans are also attracted to other humans. Note that we regard the repulsion between individuals (see Section 5.1.1) as a generalized kind of attraction between humans, i.e. a negative attraction. As the analysis of real traces suggests, encounters between social acquaintances are usually more frequent and of a longer duration compared to encounters between unrelated individuals (cf. Section 3.1).

The fundamental principle behind our model is to combine these different types of attraction: Locations exert an even greater attraction in the presence of some social acquaintances. Other locations may become only meaningful in the first place if some specific individuals are located at them. For example, a meeting room is in most cases not attractive as long as no other colleague is within the room. Thus, we let nodes move between locations (i.e. anchors), guided by the combined attraction towards locations themselves and nodes residing at them. This allows nodes to meet their social acquaintances creating long and frequent contacts between them on the level of connectivity.

Furthermore, in the context of a social relation, certain locations are of special significance that are meaningless in the context of other social relations. For example, the meeting room is significant in the context of two team members who meet there frequently. However, a team-member would hardly meet with his family in this meeting room. We show this *geo-social* relationship, i.e. the relation between social acquaintances and the social sphere, in Figure 5.2. The social acquaintances of X (center) meet at certain locations of X's social sphere (left). These locations may overlap for some social acquaintances. For example, A and B share location L as a place to meet X. Since A and B are also social acquaintances, L may represent a typical place to meet each other, too.

The idea behind the integration of geographical concepts is the following: We advocate that each mobility model should consist of two submodels: a social and a geographical mobility model. In this thesis, we only specify the former and provide certain integration points for the latter. Geographical concepts can then be specified to complement the social mobility model. For example, we do not specify how nodes move from one anchor to the next. This

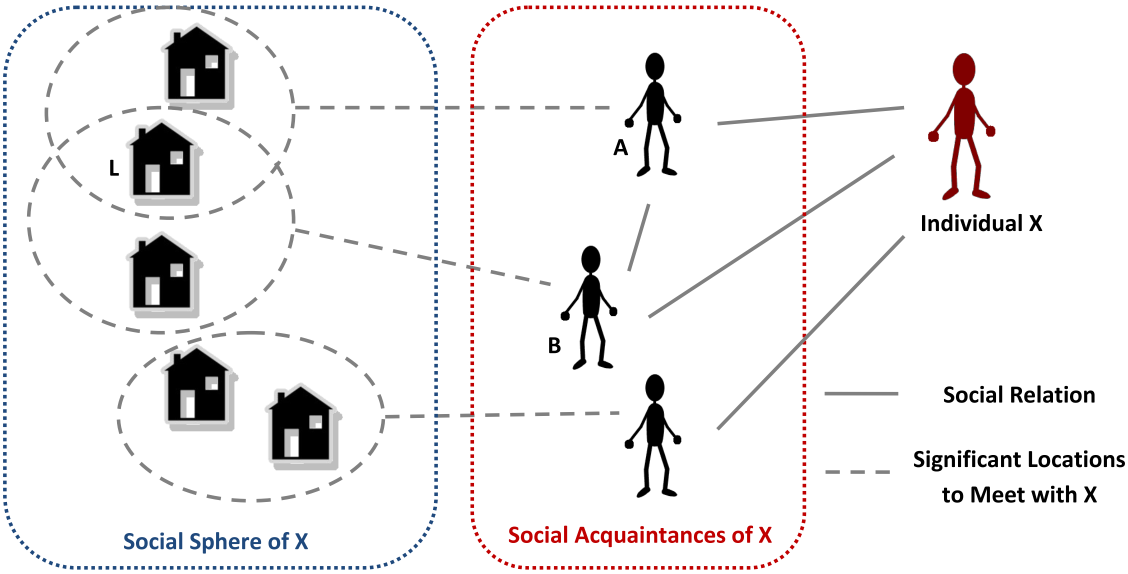


Figure 5.2: Relation between an individual's social acquaintances and social sphere.

is considered as a part of the geographical mobility model. We also propose concepts which enable the reflection of the social network despite the integration of geographical aspects.

5.2 Formal Definition of the Elements and their Dynamics

5.2.1 Elements of the Model

Let $V = \{v_1, v_2, ..., v_n\}$ denote the set of nodes and $A = \{A_1, A_2, ..., A_m\}$ with $m \geq n$ denote the set of anchors. Every node is associated with a nonempty set of *home anchors*. Nodes share their home anchors with social acquaintances as a place to meet them. The social sphere of a node may be regarded as the set of home anchors of all its social acquaintances and its own home anchors. The home anchor association function

$$\mathcal{A} : V \to \mathcal{P}(A) \setminus \emptyset,$$

where $\mathcal{P}(A)$ denotes the power set of A, maps every node to its set of home anchors. Every anchor is only home anchor of a single node, i.e. it is required that $\mathcal{A}(v) \cap \mathcal{A}(v') = \emptyset$ for all $v, v' \in V$ with $v \neq v'$. Note that by assigning multiple anchors the same coordinates on the simulation area, we still can simulate that several nodes may share a single 'home location'. Initially, all nodes are placed randomly within the geographical boundaries of one (if several exist) of their home anchors.

A simulation run is executed in discrete time steps $T = [0, 1, 2, ..., t_{max}]$. The total simulation duration t_{max} is partitioned into periodic time intervals (e.g. days) of length t_p. We call such a time interval a **simulation period**. The idea behind dividing the simulation duration into multiple simulation periods is to create temporal regularities with a periodicity of t_p.

The proposed model is based on an input social network $G = (V, E, w, \chi)$. We use weights (cf. Definition 1 on page 5), since relationships of individuals are not all of equal strength. Thus, some individuals may have a stronger influence towards each other than other individuals. This fact is of importance for mobile applications, which exploit knowledge about the social context.

Note that in addition to Definition 1, each social relation is annotated with a set of simulation ticks $\chi : E \rightarrow \mathcal{P}(T)$. If $\{u, v\} \in E$ and $t \in \chi(u, v)$, the social relation between u and v is considered to be active at time t (see Section 3.1). Hence, the *active* social context of a node v at time t can now be formally defined by

$$C_{active}(v, t) := \{u \in V \mid \{v, u\} \in E \wedge t \in \chi(u, v)\}$$

Our explicit assumption is that nodes which share a strong social relation meet more frequently and regularly than nodes which share only a weak social relation (or none). As discussed in Section 3.1, this is supported by existing work. Based on this assumption, we define the following requirement for our model:

Definition 6 (Reflection Requirement) *The number of meetings (i.e. two nodes are at the same time at the same anchor) between two arbitrary nodes $v, u \in V$ should be proportional to the strength of their relationship $w(u, v)$, for $t \rightarrow \infty$.*

If this requirement is fulfilled, we say that the input social network is **reflected** in our mobility scheme. Note that this is only one possible interpretation for the strength of the social relation. For the sake of simplicity, we use the reflection requirement as defined above. We will later show how a generalized interpretation may be integrated in a simple way (discussed in Section 8.1).

In general, our model may take an arbitrary network as input, for example extracted from real mobility traces or generated using a model for social networks like those introduced in Section 2.2.4. In Section 7.1, we will describe how we generated χ and w for the evaluation of our model.

5.2.2 Dynamics of the Model

Each node can be **associated** with an anchor. This association does not necessarily mean that a node is located at the anchor, but that it intends to reach it. In particular, a node always moves towards a so called **goal**. A goal is a point uniform randomly selected within the geographical boundaries of the associated anchor. For the sake of simplicity, we represent an anchor by a rectangular area. In general, this simplification may be easily extended to support arbitrary shapes.

The anchor association function

$$\mathcal{L} : A \times T \to V$$

specifies the set of nodes associated with an anchor $a \in A$ at time $t \in T$. In other words, we have $v \in \mathcal{L}(a, t)$ if v is either on the way to or stays at a at time t.

After the chosen goal within an anchor $a \in A$ is reached, a probabilistic **dwell time** is selected from the **dwell time distribution** D_a. D_a is a parameter of the anchor and can be modeled to capture specific target scenarios. For example, the average dwell time for a bakery may be low in comparison to the average dwell time of a theatre.

After a node v stayed with an anchor for the selected dwell time, other nodes may decide to join v on its way to the next anchor, which creates a group of nodes $G \subseteq V$. Subsequently, all nodes in G (possibly only v) choose a common **destination anchor** $d \in A$. Finally, each node in the group gets associated with d and selects a new goal randomly within d's geometrical boundaries. Finally the group begins to move towards their common destination anchor. In the following section, we illustrate the creation of groups in more detail.

5.3 Group Movements

As discussed in Section 3.5, group mobility represents an important human movement pattern that can also be exploited by mobile applications. We have integrated probabilistic group movements into our social mobility model. More precisely, if the dwell time of a node $w \in \mathcal{L}(a, t)$ expires, other nodes $v \in \mathcal{L}(a, t)$ may decide to join w for a potential group movement with probability $p_{join}(v, a, t)$.

In real life, individuals move to locations to perform some activity. As discussed in the last section, we define the time to perform such an activity by employing the anchor-specific distribution D_a. To comply with that parameter, $p_{join}(v, a, t)$ increases as the remaining dwell time of v at a decreases. Thus, nodes stay with a high probability for most of their originally chosen dwell time at the corresponding anchor. This is important since we consider the dwell time as being a parameter of the geographical mobility model.

In Figure 5.3, we show the probability that a node v joins a group movement (y-axis) dependent on v's remaining dwell time (x-axis). Δ_{dwell} denotes v's dwell time at a, selected from D_a. As shown, the probability to join increases linearly for $\Delta_{dwell} \to 0$ after v has stayed at least for a fraction $g_{thresh} \in [0, 1]$ of its original dwell time at an anchor. Thus, the

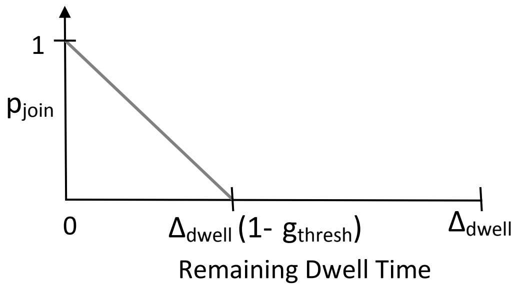


Figure 5.3: Probability to join a group, dependent on the remaining dwell time Δ_{dwell}.

probability that a node $v \in V$ at $a \in A$ joins the group movement at time $t \in T$ is defined as follows:

$$p_{join}(v, a, t) = \begin{cases} 0, \text{ if } remDwellTime(v, a, t) > \Delta_{dwell}(1 - g_{thresh}) \\ 1 - \frac{remDwellTime(v,a,t)}{\Delta_{dwell}(1-g_{thresh})}, \text{ else.} \end{cases}$$

where $remDwellTime(v, a, t)$ denotes the remaining dwell time of v at anchor a at time t.

Note that we define no micro-mobility behavior within the group because we consider this as being part of the geographical mobility model. We simply define that nodes move from the same source anchor to the same destination anchor. Based on our prototypical implementation, nodes start to move at the same time with the same constant speed. Thus, nodes are typically very close to each other, dependent on the randomly generated source and destination goals within the geometrical boundaries of the corresponding anchors.

In the following section, we illustrate how probabilistically created groups of nodes select their common destination anchor.

5.4 Probabilistic Destination Anchor Selection

A Node stays for the duration of the selected dwell time within the geometrical boundaries of an anchor. As soon as this duration expires, a new destination anchor is selected probabilistically by the node and (potentially) other nodes that join the group movement. This selection is based on different types of attraction exerted by an *anchor* towards a node. The higher the combined attraction of an anchor, the higher the probability to select this specific anchor as the next destination.

If an anchor is member of a node's social sphere, it exerts a **location attraction**. Nodes are also attracted by social acquaintances located at an anchor, that is, an anchor exerts a **node attraction**. Finally, nodes have a repelling effect on other nodes if no significant social relation exists. This is called **node respulsion** and may be regarded as a generalized type of attraction towards nodes, i.e. a negative attraction. The overall attraction of an anchor is then represented by a weighted sum over these three types of attraction. The node repulsion may diminish the node/location attraction, however, the overall attraction of an anchor cannot be negative.

We provide more details about the calculation of the individual types of attractions and their influence on the destination anchor selection in the following. The prerequisite for calculating either type of attraction towards *anchors* is to define the **social attraction** between two *nodes* $u, v \in V$ at time $t \in T$ as follows:

$$s(u,v,t) = \begin{cases} w(u,v), \text{ if } t \in \chi(u,v) \\ 0, \text{ else.} \end{cases} \tag{5.1}$$

In other words, the social attraction between two arbitrary nodes corresponds to the weight of their social relation if the relation is active. The attraction is zero if no social relation exists (thus $w(u,v) = 0$, see Definition 1 on page 5) or if the relation is not active. Note that the *social* attraction specifies the (symmetric) attraction between two individual nodes. Both the *location* and *node* attraction/repulsion, on the other hand, are defined for a node towards an anchor. In particular, the latter are dependent on the nodes which are associated with that anchor and thereby dependent on the *social* attraction towards those nodes.

This was our initial approach to calculate the social attraction between two nodes. Later, we will introduce an advanced formula for calculating the social attraction. This advanced formula may replace Equation (5.1) to yield a more robust reflection of the social network in the presence of some 'disturbing' factors, as we will explain in Section 5.5 in more detail. However, this results in additional complexity. Thus, we consider the advanced calculation of the social attraction as an (optional) extension of the basic model.

5.4.1 Location Attraction

The attraction an anchor exerts towards a node is dependent on the node and the current simulation tick. To support the time-dependent periodic location attraction, every anchor $a \in A$ is associated with a characteristic **anchor function**

$$f_a : T \to [0,1].$$

f_a is a t_p-periodic function, i.e.

$$f_a(t) = f_a(t + t_p) \text{ for all } t \in T \ : \ t + t_p \leq t_{max}.$$

The basic idea behind employing a periodic anchor function rather than a constant attraction is to introduce a temporal dependency (see Section 3.4). This dependency is periodic which results in the habitual behavior of nodes with periodicity t_p, as we will show in the evaluation of our model.

In general, the shape of the anchor function may be arbitrary. However, to actually yield habitual behavior, the anchor function should not be constant. In our implementation of the social mobility model, the anchor function is randomly generated for each anchor using different simple models. Two examples are depicted in Figure 5.4. Two functions, one with a periodic peak (black) and one with a periodic plateau (grey), are depicted. The peak-function may be a good representation for a canteen while the plateau-function may rather be used to represent an office building. We generate the position of the peak/plateau randomly for each anchor.

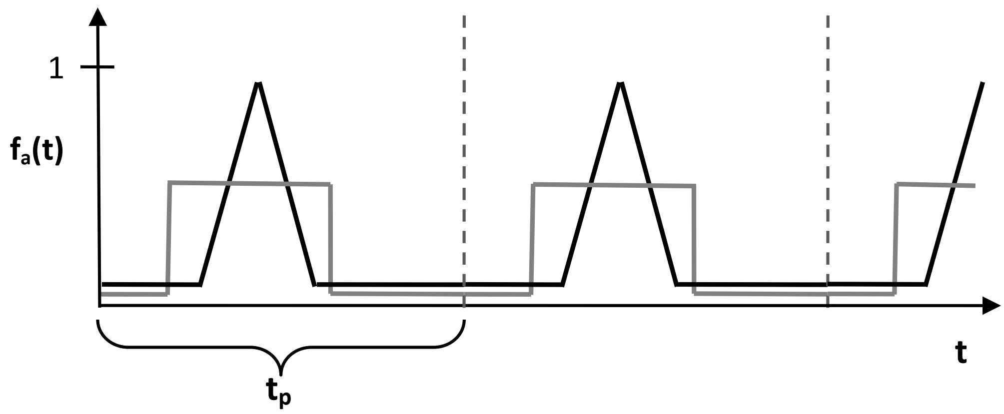


Figure 5.4: Example for two characteristic anchor functions.

In order to yield spatial regularities (cf. Section 3.3), each node is only affected by the location attraction of a subset of all anchors. This subset corresponds to the social sphere of this node. Basically, nodes share the location attraction of their home anchor(s) with all members of their active social context. This reflects the real world where, for example, we seldom let strangers in our house. For each node v, the influence of the location attraction exerted by the home anchor of another node u is proportional to the social attraction between u and v. Thus, the probability to spend time at a foreign anchor and therefore possibly with the associated node reflects the strength of their relationship. In particular, this complies with the reflection requirement.

Based on this discussion, the location attraction an anchor $a \in A$ exerts towards a node $v \in V$ is defined as follows:

$$A_{loc}(v, a, t) := \begin{cases} s(v, v_h, t) \cdot f_a(t), \text{ if } v \neq v_h \\ \gamma \cdot f_a(t), \text{ else.} \end{cases} \tag{5.2}$$

where a is a home anchor of v_h, i.e. $a \in \mathcal{A}(v_h)$, and $\gamma \in [0, 1]$ is a constant factor to avoid that the attraction towards the home anchor(s) dominates the attraction towards other anchors. γ allows to control the time a node spends at its own home anchor(s). We used $\gamma = 1$ for our basic model. However, we will later introduce an advanced calculation of the social attraction which makes it necessary to use a $\gamma < 1$.

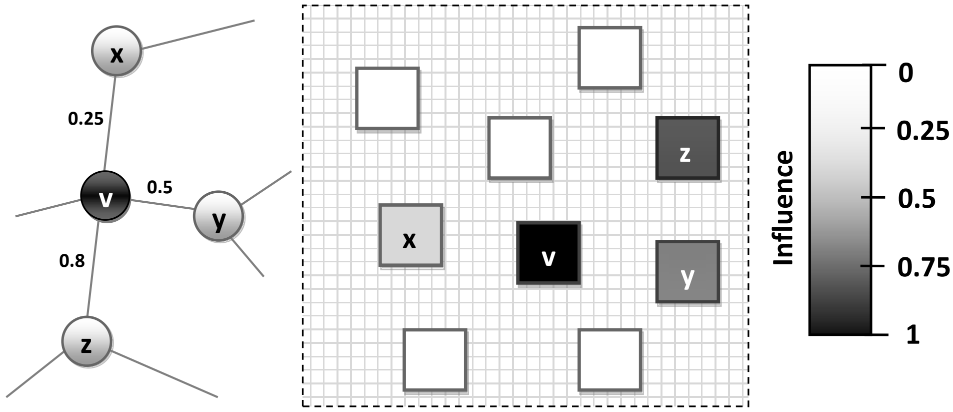


Figure 5.5: Example for the social sphere of node v.

Figure 5.5 illustrates the social sphere of a node v, created by the location attraction. On the left, we show an excerpt of a social network. In the center, we depict the simulation area with several anchors. Home anchors of nodes shown in the excerpt of the social network are captioned with the associated node. The darker we have depicted the color of an anchor, the greater its influence on v.

5.4.2 Node Attraction

An anchor a exerts a node attraction towards a node v if at least one node of v's active social context is located at a. The strength of this attraction, however, depends on the social attraction towards *all* nodes which are associated with a.

First, we propose a simple formula (A^I_{node}) for the calculation of the node attraction. Then, we discuss some difficulties that may arise from the integration of a geographical mobility model. As a consequence, the existing formula is refined which yields an alternative formula

(A^{II}_{node}) that enables better results if geographical concepts are integrated, but at the price of requiring additional computing power.

Simple Calculation

To satisfy the reflection requirement, the probability to produce a meeting between two nodes must be proportional to the strength of their social relation. At the same time, the presence of nodes to which no social relation exists should lead to a rather weak attraction. Hence, the node attraction an anchor $a \in A$ exerts towards a node $v \in V$ at time $t \in T$, is calculated as follows:

$$A^{I}_{node}(v, a, t) := \frac{1}{|\mathcal{L}(a, t)|} \sum_{u \in \mathcal{L}(a,t)} s(v, u, t). \tag{5.3}$$

In other words, the node attraction is proportional to the mean social attraction towards all nodes which are associated with a at time t.

Enhanced Calculation

In the presence of geographical concepts, some nodes may require a long time to travel between anchors. For example, obstacles and streets may constrain the movement of nodes which leads to long travel times. Long distances or heterogeneous speed distributions may create a similar effect. A specific behavior can be observed under such conditions: A node v may be attracted by another node u, staying currently at an anchor a. By the time v reaches a, u is already on its way to another anchor. Hence, v misses the opportunity to meet u, though this was the reason to visit a in the first place (*Case A*). The greater the time to travel between anchors compared to the average dwell time, the more frequently this behavior can be observed.

In addition to case A, we have identified another similar case. Consider the following situation: u is associated with an anchor a, but has not reached it. v may be attracted to a based on the node attraction towards u. We assume that u requires a lot of time to reach a. Thus, by the time u finally reaches a, v has already chosen a new destination anchor. Hence, they miss the opportunity to meet each other (*Case B*).

To make the integration of geographical mobility models easy, nodes should be able to meet each other even in the presence of long travel times. Therefore, we introduce an enhanced formula to calculate the node attraction. This formula is quite similar to Equation (5.3), but differs in the set of nodes considered in the calculation. More precisely, a node is only considered in the calculation if a meeting will (presumably) occur.

In the following, we discuss the example shown in Figure 5.6. Node u considers anchor a as a potential destination anchor. Thus, the node attraction of a towards u has to be

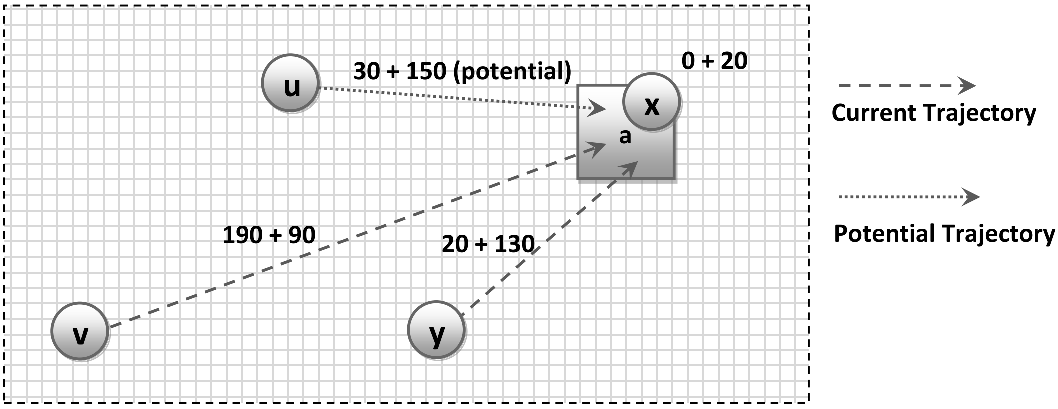


Figure 5.6: Example Scenario: Node u considers anchor a as a potentiell destination. Travel and dwell times are shown above the arrows. In this scenario, only node y is considered for the calculation of the social attraction.

calculated. We assume that u would require 30 simulation ticks to reach a. A potential dwell time of 150 simulation ticks is selected from D_a. If u actually chooses a as the new destination anchor later, this potential dwell time will become the actual dwell time for the next stay. Three nodes are currently associated with a. Node x is already located at a and will choose a new destination anchor in 20 simulation ticks. Hence, no meeting between u and x will occur (case A) and x is not considered in the calculation of the social attraction. Nodes v and y are both on their way to reach a. v will reach a not until 190 simulation ticks later. At this time, u will already have chosen a new destination anchor (case B), thus v is not considered either in the calculation of the social attraction. The only node considered is y because a meeting with u will presumably occur 30 simulation ticks later.

To determine if two nodes will meet at an anchor, it is required to choose a potential dwell time at the time of the selection of the new destination anchor. This way, the entire time window a node will stay at an anchor is known. Hence, a node may determine if a meeting with other associated nodes will occur. We denote the remaining time it takes v to reach a at time t as $travelTime(v, a, t)$. In particular, if v has already reached its goal within the geometrical boundaries of a at time t, we have $travelTime(v, a, t) = 0$. Hence, we propose the following enhanced formula as an alternative for Equation (5.3):

$$A^{II}_{node}(v, a, t) := \frac{1}{w} \sum_{v' \in \hat{\mathcal{L}}(a,t')} s(v, v', t)$$
$$\text{with } w = |\hat{\mathcal{L}}(a, t')| \text{ and } t' = t + travelTime(v, a, t),$$

where $\hat{\mathcal{L}}(a,t')$ is an approximation of $\mathcal{L}(a,t')$ at the future time $t' > t$. Note that the only difference compared to Equation (5.3) is the set of nodes considered in the sum. To combine the cases A and B, we define that $u \in \hat{\mathcal{L}}(a, t + travelTime(v,a,t))$ if and only if

$$travelTime(u,a,t) + remDwellTime(u,a,t) > travelTime(v,a,t) \quad \text{(Case A)}$$
$$\wedge\ travelTime(u,a,t) < travelTime(v,a,t) + \Delta_{dwell}^{pot} \quad \text{(Case B)}$$

where Δ_{dwell}^{pot} denotes the potential dwell time of v in the case that v chooses a as the new destination anchor. Just like in the case of the simple calculation, Δ_{dwell}^{pot} is selected from the distribution D_a. Another approach may be to choose Δ_{dwell}^{pot} in a way that v meets as many nodes as possible. However, this would constrain the geographical mobility model and the ability to parameterize the simulation for different scenarios.

Note that at a time t'' with $t' > t'' > t$, another node may decide to select anchor a as a new destination anchor. Or, a node $u \in \hat{\mathcal{L}}(a,t')$ may leave a before its dwell time expired because u joined a group movement. Thus, $\hat{\mathcal{L}}(a,t')$ is only an approximation of the actual $\mathcal{L}(a,t')$ at time t'.

The illustrated enhanced calculation of the node attraction causes additional costs in terms of computing power. In the following, these costs are enumerated, based on the scenario that an arbitrary node v chooses a new destination anchor.

- For every *potential* destination anchor, i.e. for all $a \in A$ with $\exists u \in \mathcal{L}(a,t) : u \in C_{active}(v,t)$, a potential dwell time has to be selected from the distribution D_a. If the generation of values according to D_a is simple, like in the case of a Gaussian or uniform distribution, these costs are negligible.
- For every potential destination anchor, the potential travel route and time have to be calculated. This is strongly dependent on the geographical mobility model. However, many mobility models use a graph to represent pathways. In this case, efficient algorithms exist which calculate the shortest path between two arbitrary locations.
- For every node, the remaining travel time and dwell time have to be known at any point in time. But even if the simple approach is used, the route to the next destination and the dwell time have to be calculated at some point. Thus, additional overhead is only created by maintaining and decrementing these values, which is negligible compared to the overall complexity of our model.

Thus, in typical cases, the enhanced calculation increases the costs only slightly.

5.4.3 Node Repulsion

To satisfy the reflection requirement, pairs of nodes sharing a weak social relation and pairs of nodes without any social relation should meet with low probability. However, in experiments we have observed that this is not always the case. More precisely, we have implemented the destination anchor selection based only on the location and node attraction as discussed so far. We have discovered that certain pairs of nodes meet frequently despite sharing only a weak social relation or no social relation at all. In particular, this happens if both nodes share a common social acquaintance v. The reason for this behavior is that both nodes are *indirectly* attracted by v and v's home anchor(s). The general case of multiple common social acquaintances is shown in Figure 5.7. Two nodes without any social relation (u and v) are depicted, which are strongly attracted to each other due to multiple common social acquaintances. The more common social acquaintances two nodes have, the more pronounced is this behavior.

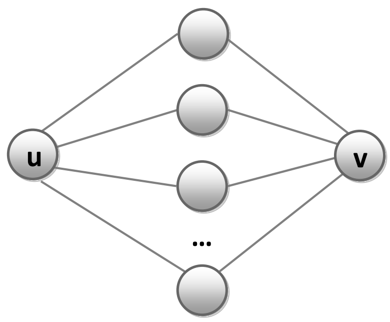


Figure 5.7: Typical scenario in which a node repulsion is necessary.

To prevent meetings between nodes based only on the attraction towards common social acquaintances, we introduce a node repulsion. The reader may argue that it might be a realistic human behavior that two individuals sharing a common friend also share a social relation and meet frequently. Indeed, as discussed in Section 2.2.3, this is a common characteristic of social networks. However, it is not the goal of our mobility model to make any assumptions about how relationships between individuals are structured (cf. Section 1.2). Such assumptions can be integrated into our mobility model by providing an appropriate social network as input. Furthermore, in the evaluation of our model (Section 7.3), we will show that the node repulsion also plays a vital role for the emergence of realistic characteristics of the inter-contact distribution.

First, we assume that the simple formula for the social attraction is employed and illustrate the calculation of the node repulsion in this case. Afterwards, we will explain how to change the calculation if the enhanced formula is used.

The further discussion is based on the following scenario: At time t, a node v considers an anchor a as a potential next destination. A node v_{weak} is associated with a at time t, i.e. $v_{weak} \in \mathcal{L}(a,t)$. We assume that either no social relation between v and v_{weak} exists, the social relation is not active, or that the weight of the relation is below the so called **penalty threshold** $w_{weak} > 0$. In other words, we assume $\exists v_{weak} \in \mathcal{L}(a,t) : s(v, v_{weak}, t) < w_{weak}$.

Of course, two nodes sharing a weak relation should also meet on occasion. We only want to avoid meetings between nodes based on the attraction towards a common social acquaintance. Therefore, the node repulsion is only considered in two cases:

1. In addition to v_{weak}, at least one node which shares a strong, active social relation with v stays at a, i.e.

$$\exists v_{strong} \in \mathcal{L}(a,t) : s(v, v_{strong}, t) \geq w_{weak}.$$

   This represents the case in which v and v_{weak} are attracted not by each other, but by their common social acquaintance v_{strong}. Although we make no assumption on the relationship between v_{strong} and v_{weak}, their co-location implies that they share a significant social relation with high probability.

2. a is neither home anchor of v, nor home anchor of v_{weak}, i.e.

$$a \notin \mathcal{A}(v) \wedge a \notin \mathcal{A}(v_{weak}).$$

   This represents the case in which the attraction between v and v_{weak} is based on their common attraction towards a home anchor of a common social acquaintance.

If at least one of the above described cases is satisfied, we calculate the node repulsion as follows:

$$A_{rep}(v,a,t) := -\frac{0.5}{w_{weak}} \cdot \sum_{v' \in \mathcal{L}(a,t))} max(w_{weak} - s(v,v',t), 0).$$

If neither case is satisfied, we define $A_{rep}(v,a,t) = 0$. In other words, each node v_{weak} with $s(v, v_{weak}, t) < w_{weak}$ creates a penalty inversely proportional to the weight of the social relation, and all such penalties are accumulated to yield the node repulsion. We have employed the fraction $\frac{0.5}{w_{weak}}$ such that a node u with $w(u,v) = 0$ always creates a penalty of 0.5[1], independent of the definition of a 'weak' social relation (i.e. the penalty threshold w_{weak}).

If the enhanced formula for the social attraction is used, $\mathcal{L}(a,t)$ is simply replaced by $\hat{\mathcal{L}}(a,t')$ with $t' = t + travelTime(v,a,t)$ for each occurrence.

[1]Note that the value of 0.5 is completely arbitrary. A change of this value only changes the optimal value of the parameter ϕ that weightens the influence of the node repulsion, which we will introduce in the following.

5.4.4 Selection of the Destination Anchor

After a node's dwell time has expired, a new destination anchor is chosen by the node or group of nodes. This probabilistic selection is based on the discussed types of attraction. We use a weighted sum to represent an anchor's overall attraction.

Let $G \subseteq V$ be the group created by the algorithm described in Section 5.3. We note that in particular, G may contain only a single node. The overall attraction an anchor a exerts towards a group of nodes G at time t, is calculated as follows:

$$A_{overall}(G,a,t) := max(0, \sum_{v \in G} [\alpha A_{node}(v,a,t) + (1-\alpha)A_{loc}(v,a,t) + \phi A_{rep}(v,a,t)])^g$$

with $\phi, \alpha \in [0,1]$, $g \in [0, \infty[$, and $A_{node} \in \{A^I_{node}, A^{II}_{node}\}$. In other words, we accumulate the attraction an anchor exerts towards the individual nodes in G.

g controls the 'greediness' of the destination anchor selection. $g \to 0$ implicates a completely random selection while $g \to \infty$ represents the case of a deterministic selection of the anchor that exerts the highest overall attraction.

α controls the weight between location and node attraction. For $\alpha \to 0$, a node chooses its destination primarily based on the current value of the anchor function. Thus, the movement of nodes is strongly dictated by habitual behavior. For $\alpha \to 1$, on the other hand, the movement of nodes is mostly based on the current location of social acquaintances. Hence, α controls the degree of randomness.

ϕ, which we call the **penalty factor**, controls the strength of the node repulsion. A low ϕ may lead to a poor reflection for some pairs of nodes. A high ϕ may lead to less meetings between nodes.

After the calculation of the overall attraction of all anchors, we can finally define the probability that a group of nodes G chooses an anchor a as the next destination anchor at time t:

$$P(\text{Next Destination Anchor} = a) = \frac{A_{overall}(G,a,t)}{\sum_{a' \in A} A_{overall}(G,a',t)}$$

Hence, the probability of choosing an anchor as the next destination is proportional to its overall attraction towards the group.

5.5 Robust Reflection of the Social Network

The probabilistic selection of the destination anchor as discussed so far results in a good overall reflection, i.e. averaged over all social relations, as we will show in the evaluation of our model (Chapter 7). However, we observed a rather poor reflection for some specific social relations. Either considerable too many or too few contacts were produced between such nodes with respect to their social relation. First, we analyze and identify the reason for this behavior (Section 5.5.1). In Section 5.5.2, we give an overview of our generic solution. Afterwards, we discuss the solution in more detail (Sections 5.5.3 and 5.5.4). We consider the results in this section to be an extension of the basic model proposed so far. The concepts can be integrated without much change. They improve the reflection of the social network at the cost of additional complexity.

5.5.1 Motivation for further Improvement

The poor reflection of some social relations is caused by three different factors. We discuss these factors in the following.

Structural Properties of the Social Network

In our experiments we found a correlation between the probability that two nodes meet and their similarity, i.e. the degree of common social acquaintances (cf. Section 2.3.1). We use an example shown in Figure 5.8 to illustrate the reason for this correlation. A simple excerpt of a social network is depicted. All existing edges between the four illustrated nodes are shown and their weights are assumed to be equal. We observe that nodes A, B, C and D form a clique. Node E has only a relation to a single member of the clique (node A). Every node is assumed to have a single home anchor. Thus, the social sphere of A includes the home anchors of A,B,C,D and E. Four of these anchors are also part of the social sphere of B. Thus, A may encounter B on any of these four anchors. However, there are only two anchors at which A may meet E. In addition, A is able to meet with C or D and at the *same time* meet with B since all are social acquaintances of each other. Thus, based on the probabilistic destination selection described in Section 5.4, the probability to meet E is smaller than the probability to meet B. This problem occurs only because A shares some (in this case two) common social acquaintances with B, but none with E.

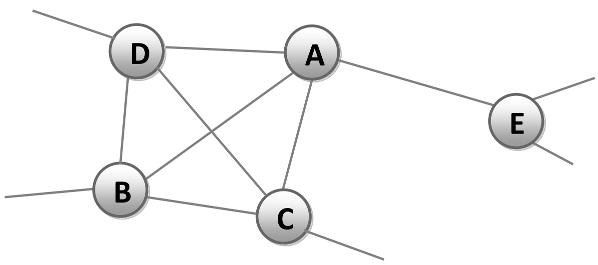


Figure 5.8: The strong similarity between A and B leads to an increased meeting probability.

As illustrated by this example, the general problem is that two social acquaintances sharing an above-average similarity meet with increased probability because they may meet at a larger number of anchors compared to social acquaintances which share a rather low similarity. Additionally, meetings between multiple nodes at the same time may be produced within strongly clustered subgraphs. Thus, the reflection of the social network may be degraded due to structural properties.

Concepts of the Geographical Mobility Model

The integration of concepts of the geographical mobility model may cause a poor reflection of some social relations.

We consider the case that node u has a social relation of equal weight with the nodes x and y. We assume that x requires a above-average time to travel between anchors because of long distances, low speed, or obstacles. y on the other hand, travels very fast between anchors. Thus, u may produce much more meetings with x than with y, despite the fact that both social relations have the same weight.

We assume that there are many other geographical concepts which may lead to a similar disruption of the reflection of the social network.

Heterogenous Sociability

In a typical social network, the node degree is very heterogeneous. Some nodes have many more social relations than other nodes. Thus, to satisfy the reflection requirement, a node which has many social relations, e.g. a hub in a scale-free network, has to produce much more meetings per time with its social acquaintances compared to a node which has only few social relations. This may lead to a poor reflection between pairs of nodes characterized by a very different node degree.

We illustrate this problem by considering the example social network shown in Figure 5.9. For the sake of simplicity, we assume that all social relations have a weight of 1 and are active during the whole simulation duration. As depicted, node B has only a single social acquaintance and may, so to speak, dedicate all its time to meet node A. Node C, however, has many social acquaintances and may have only a limited time to meet with A. Thus, A may produce much more meetings with B than with C, although both social relations have an equal strength.

We have not considered so far that social relations may have different weights. Furthermore, social relations are not active all the time. Thus, we propose a measure for the number of

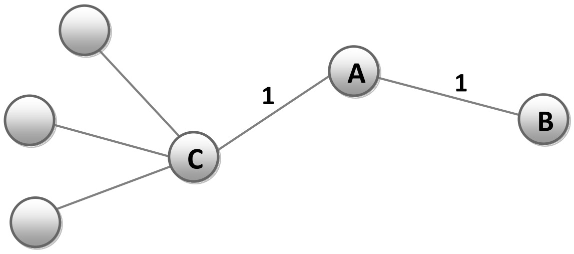


Figure 5.9: A Social network characterized by an asymmetric node degree.

weighted social relations at a given time during the simulation. We define the **sociability** of a node $v \in V$ at time $t \in T$ as follows:

$$soc(v,t) := \sum_{x \in C_{active}(v,t)} w(v,x)$$

In other words, the sociability measures the sum of the weights of all *active* social acquaintances at a given time. To avoid the above described problem related to the heterogeneous sociability, we introduce the following definition.

Definition 7 (Sociability Requirement) *The sociability of a node must be proportional to the number of produced meetings per time.*

The real-world intuition behind this is that individuals characterized by many social relations typically meet more frequently with other individuals. If Definition 7 is satisfied, all nodes produce the same number of meetings for social relations of comparable weight. This is required to satisfy the reflection property as we have demonstrated in the example above. Unfortunately, this is not the case based on the basic model we have proposed so far. Indeed, measurements in experiments show that the number of produced meetings per time unit correlates to the average sociability of a node. However, this correlation is significantly below a proportional relationship. Thus, we require concepts to ensure that the sociability requirement is satisfied.

5.5.2 Overview

In the following, we give an overview of the concepts to solve the discussed problems and show how they are related.

Isolation Phase

The idea to satisfy the sociability requirement is the following: Individuals characterized by few social relations typically spend more of their time in isolation compared to hubs, which are frequently surrounded by social acquaintances. To simulate this typical behavior, we introduce a so-called **isolation phase**. If a node v selects a new destination anchor, it may *probabilistically* enter into an isolation phase. In this case, v moves to a randomly chosen point on the map that is outside of the boundaries of any anchor. This point is called the **isolation location**. v stays at the isolation location for a specific time interval. During this time interval, no other node is attracted by v. Thus, a node in an isolation phase produces no meetings with other nodes. After this duration expires, v may select a new destination anchor again and thus leave the isolation phase. Or, v may enter into an isolation phase again. The isolation phase may be regarded as a regular visit of an anchor without producing any explicit meetings with other nodes.

In our prototypical implementation, the isolation locations are randomly selected on the map. On the level of connectivity, the geographic position is of no importance because no connectivity is supposed to be created during an isolation phase. The use of an isolation location is just the means to force a node to be isolated for some time. If a mobile application explicitly requires geographical data such as GPS coordinates, the isolation locations have to be placed according to a scenario-dependent, realistic scheme.

The challenge to satisfy the sociability requirement lies in the calculation of the **isolation probability** $p_{iso}(v,t)$ which specifies the probability that a node $v \in V$ enters into an isolation phase if its current dwell time expires at time $t \in T$. As discussed above, the goal is to create a proportional relation between the sociability and the number of meetings per time. Thus, $p_{iso}(v,t)$ must dependent both on v's current sociability and the number of meetings between v and active social acquaintances $u \in C_{active}(v,t)$. For example, the isolation probability of node B in Figure 5.9 is usually higher than the isolation probability of node C.

Correction Factor

As discussed above, some structural properties and geographical concepts may create disruptions in the reflection of the input social network. Some nodes may have too few or too many encounters, compared to the weight of their social relation. Concepts like group movement may also create such disruptions. We propose the following as a generic solution: At periodic time intervals, the number of meetings between all pairs of social acquaintances is compared to the weight of the corresponding social relation. If too many (few) meetings between two nodes have been produced, their social attraction is decreased (increased). By employing this feedback control loop in the presence of disruptions, eventually a stable equilibrium (as defined by the reflection requirement) is reached again. This feedback control loop is shown in Figure 5.10.

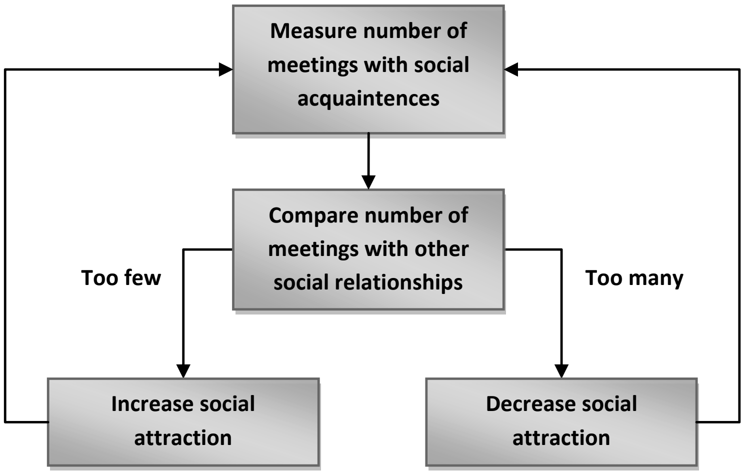


Figure 5.10: Feedback control loop to update the social attraction between nodes.

To adjust the social attraction between two nodes, we introduce the so-called **correction factor**. This numeric value is multiplied with the weight of the social relation to yield the advanced social attraction. Thus, we replace Equation (5.1) on page 46 by the following formula:

$$s(u,v,t) = \begin{cases} c(u,v,t) \cdot w(u,v), \text{ if } \{u,v\} \in E \wedge t \in \chi(u,v) \\ 0, \text{ else.} \end{cases}$$

where $c(u,v,t) \in [0,1]$ denotes the correction factor at time t for the social relation between u and v. At the beginning of the simulation, the correction factor is initialized to 1 for all social acquaintances. If too many (too few) meetings between to nodes have taken place, for instance because of above described disruptions, d decreases (increases) towards 0 (1). Thus, the probability to meet decreases (increases) and eventually the number of meetings approaches the desired value with respect to other social relations and the corresponding weights. We used a constant factor of $\gamma = 0.4$ for the calculation of the location attraction (Equation (5.2) on page 48) to avoid that a node spends all its time at its home anchor because, as we will become apparent in the following, on average we have $c(u,v,t) \approx 0.5$. This value turned out to yield better results.

Note that there is also a real-world intuition behind the correction factor: Real people also try to balance their social activities based on the strengths of their social relationships. For example, people remember that they have not met a certain social acquaintance lately although they value this relationship. Thus, they make plans to meet this individual more frequently in the near future.

A Combined Approach

The idea behind calculating both the correction factor and the isolation probability is the following: During the simulation, we maintain the number of meetings between each pair of social acquaintances. Periodically, it is determined how many meetings should have been produced if the input social network was reflected *perfectly*, based on the sum of all meetings a node has produced. We call the number of produced meetings between two nodes $u, v \in V$ in the theoretical case of a perfect reflection the **meeting-quota** between u and v. To yield the meeting quota, essentially, all produced meetings of a node (at the time of the calculation) are *redistributed* as defined by the reflection requirement.

Based on a theoretical/actual comparison, it is possible to calculate the correction factor which implements the described feedback control loop.

In general, the sum of all meeting-quotas for a node may be less than the total sum of produced meetings for this node. This is a typical sign that the node has produced too many meetings, based on the sociability requirement. The difference between the sum of meeting-quotas and the total sum of meetings is then used to calculate the isolation probability.

5.5.3 Calculation of the Meeting-Quota

We employ a simple example to motivate the algorithm behind the calculation of the meeting-quotas. Afterwards, we present the algorithm in a formal way.

Motivational Example

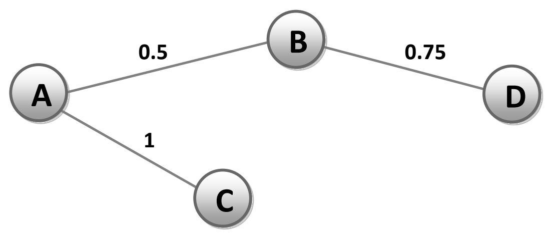


Figure 5.11: Example social network to illustrate the calculation of the meeting-quotas.

We consider the example social network shown in Figure 5.11. All depicted social relations are assumed to be active. In Table 5.1, we show the sociability and the total number of

Node	Sociability	Total Number of Meetings	Social Influence
A	1.5	150	100
B	1.25	140	112
C	1.0	120	120
D	0.75	100	125

Table 5.1: Different quantities at time t_i for the example social network shown in Figure 5.11

meetings for each node at some time $t_i \in T$. Though constructed, this example represents typical situations we have observed in our experiments.

Node A has a social relation with a weight of 0.5 with node B and a social relation of weight 1 with node C. Thus, the perfect reflection of the social network would yield twice as much meetings with C than with B. Hence, to satisfy this requirement, A should have produced 100 meetings with C and 50 meetings with B at time t_i. However, this calculation is based on the perspective of A. The same calculation, based on the perspective of B yields different results. In particular, one could argue that B should have produced $140 \cdot (0.5)/(0.5 + 0.75) = 56$ meetings with A at time t_i. This contradicts the value that we have calculated from A's perspective (50 meetings). However, both A and B should adapt to the same quantity. Otherwise, B would try to produce more meetings with A while A would 'avoid' to meet B. This is no desired behavior. Thus, we consider it a requirement to calculate *symmetric* meeting-quotas.

To gain symmetric quota values, we identify a unique quantity, which assigns a symmetric meeting-quota to each social relation dependent on the weight. However, as discussed, some nodes may produce more meetings per time and social relation than others. At his point, the calculation of the isolation probability comes in. Based on the sum of meetings for each node, B is able to dedicate more meetings per time than A into a social relation of a fixed weight, as we have calculated. This may be measured by the following fraction, which we call the **social influence** of a node:

$$\text{Social Influence} = \frac{\text{Total Number of Meetings}}{\text{Sociability}}.$$

We will later give a more formal definition. A node with a below-average social influence (such as A) may have difficulties to reflect its social relations properly. On the other hand, a node with an above-average social influence (such as D) may produce too many meetings with its social acquaintances with respect to the weight of the social relation. This results in the violation of the reflection requirement (cf. Section 5.5.1). Assume that the social influence is equal among all nodes. Then, by definition, for each node v, the total number of meetings with social acquaintances is proportional to the sociability of v, thus satisfying the sociability requirement. Hence, our goal is to adjust the isolation probability to yield a constant social influence among all nodes.

In our example (see Table 5.1), the social influence of A is smaller than the social influence of B. Thus, we let B enter into an isolation phase more often than A. This decreases the social influence deviation between A and B and eventually leads to a balance between those nodes.

In general, we identify the node with the smallest social influence. We call the social influence of this node the **reference social influence**, which corresponds to the unique quantity we have mentioned above. All other nodes adjust to the reference social influence by entering probabilistically into an isolation phase, dependent on how much their social influence differs from the reference social influence. We use the *smallest* social influence since it is always possible to force a node to produce less meetings (due to isolation phases) but not vice versa. In our example (see Table 5.1), the reference social influence corresponds to 100.

The *difference* between the total sum of meetings and the sum of meeting-quotas may be used as a measure for the intended number of isolation phases. The higher the difference, the stronger the deviation from the reference social influence. In the following, we will define the concepts behind the isolation probability and the correction factor in a more formal way.

Formal Definition of the Algorithm

For each node, the number of meetings with every other node is maintained. The function

$$m : V \times V \times T \to \mathbb{N}$$

keeps track of the number of meetings between two nodes at a given simulation tick. In particular, m is symmetric in the first two arguments. For convenience, we define

$$M(v,t) := \sum_{u \in C_{active}(v,t)} m(v,u,t) \tag{5.4}$$

as the total number of meetings a node v has produced *with its active social acquaintances* at time t.

In the above described example, all social relations where assumed to be active. In general, some relations may not be active and should therefore not be included in the calculation because no meeting between such nodes should be produced at this time. Hence, we define the social influence at time $t \in T$ formally as follows:

$$i(v,t) := \frac{M(v,t)}{soc(v,t)}.$$

We assume that at least one social relation is active at any given time for each node. At $t \in \{t_{update}, 2t_{update}, 3t_{update}, \ldots\}$, for each pair of social acquaintance $\{u,v\} \in E$, we calculate the **meeting-quota** $quota(v,u,t)$, which represents the number of meetings between u and v in the theoretical case of a *perfect* reflection at time t, as defined by the reflection requirement. The meeting-quota must satisfy the following *requirements*:

- **(Requirement 1)** Every meeting-quota for *active* social acquaintances has to be proportional to the weight of the relation, i.e.

$$\forall v \in V,\ \forall x, y \in C_{active}(v,t) : \frac{quota(v,x,t)}{quota(v,y,t)} = \frac{w(v,x)}{w(v,y)} \tag{5.5}$$

  This is required to satisfy the reflection requirement.

- **(Requirement 2)** The meeting-quota for two social acquaintances must be symmetric in the first two arguments, i.e. both nodes must have the same notion on how many meetings they should have produced in the case of a perfect reflection at time t. Thus, we require that

$$\forall \{u,v\} \in E :\ quota(u,v,t) = quota(v,u,t).$$

  We have discussed the reasons for this requirement in the motivational example.

- **(Requirement 3)** For each node, the sum of all meeting-quotas should be less than or equal the sum of actual meetings, i.e.

$$\forall v \in V :\ M(v,t) \geq \sum_{u \in C_{active}(v,t)} quota(v,u,t).$$

  If this would *not* be the case, a node may be unable to produce enough meetings per time to reach the meeting-quota.

After setting up the requirements, we can illustrate the calculation of the meeting-quota. First, we define the above discussed reference social influence formally as follows:

$$q(t) := \min_{v \in V} i(v,t)$$

In particular, $q(t)$ represents the meeting-quota for social acquaintances having a weight of 1. To assign a meeting-quota to arbitrary social relations $\{u,v\} \in E$, we use the following formula:

$$quota(v,u,t) := \begin{cases} w(v,u) \cdot q(t), & \text{if } t \in \chi(u,v) \\ 0, & \text{else.} \end{cases} \tag{5.6}$$

In particular, the meeting-quota for two nodes whose relation is not active is zero. Based on the symmetry of $m(.,.,t)$ and $w(.,.)$, the meeting-quota is symmetric as well (Requirement 2). By putting Equation (5.6) into Equation (5.5), it can be easily verified that Requirement 1 is satisfied. Since we defined $q(t)$ as the minimum social influence, we have $M(v,t) \geq \sum_{u \in V} quota(v,u,t)$ for all $v \in V$ which satisfies Requirement 3.

In our motivational example (Figure 5.11 and Table 5.1), node A corresponds to the node with the smallest social influence, i.e. $q(t_i) = 100$. Thus, it follows that $quota(A,C,t_i) = 100$, $quota(A,B,t_i) = 50$, and $quota(B,D,t_i) = 75$.

5.5.4 Calculation of the Correction Factor and the Isolation Probability

As discussed, the greater the difference between the total number of meetings and the sum of assigned quota values, the higher the deviation from the reference social influence. We use a probability to enter into an isolation phase which is proportional to this difference. Thus, the probability that a node v enters into an isolation phase if its dwell time expires at time t is calculated as follows:

$$p_{iso}(v,t) := \frac{M(v,t) - \sum_{u \in V} quota(v,u,t)}{M(v,t)}. \tag{5.7}$$

Due to Requirement 3, we have $p_{iso}(v,t) \in [0,1]$ for all $v \in V, t \in T$. The calculation based on our motivational example yields $p_{iso}(A,t_i) = 0$, $p_{iso}(B,t_i) = 15/140 \approx 0.10$, $p_{iso}(C,t_i) = 20/120 \approx 0.167$ and $p_{iso}(D,t_i) = 25/100 = 0.25$.

To adjust the correction factor for two social acquaintances $\{v,u\} \in E$, the meeting-quota has to be compared to the actual number of meetings at $t \in T$. Hence, we determine the following deviation from the meeting-quota

$$\delta(v,u,t) := quota(v,u,t) - m(v,u,t).$$

If too few meetings between two nodes $u, v \in V$ were produced at time t, we have $\delta(u,v,t) > 0$. Respectively, if too many meetings were produced, we have $\delta(u,v,t) < 0$. $\delta(u,v,t) = 0$ represents the ideal case of a perfect reflection.

To obtain a correction factor $c(v,u,t) \in [0,1]$, we map all deviations to the interval $[0,1]$ as follows:

$$c(v,u,t) := \frac{\delta(v,u,t) - \delta_{min}}{\delta_{max} - \delta_{min}}$$

with $\delta_{min} = \min(I)$, $\delta_{max} = \max(I)$ where $I = \{\delta(v,x,t) | x \in C_{active}(v,t)\} \cup \{\delta(u,x,t) | x \in C_{active}(u,t)\}$. Note that this is simply a linear transformation from I to $[0,1]$. We chose I in this way to obtain *symmetric* correction factors that are at the same time distributed over the whole interval $[0,1]$.

Figure 5.12 shows a typical example for the behavior of the social attraction between two social acquaintances u, v with $w(u,v) = 1$. This example was measured in an experiment. We show both the social attraction (solid black) and the deviaton δ (dashed grey) on the y-axis against simulated time (x-axis). We display the exact value of the social attraction, while the deviaton was scaled down to fit the plot. As defined, the social attraction starts out with a value of 1. For $t < 40000$, too few meetings between v and u have been produced (deviation > 0), thus $s(u,v,t)$ stays close to 1. This leads to frequent meetings between u and v, and hence towards a perfect reflection (deviation $\rightarrow 0$). After $t = 40000$, eventually too many meetings have been produced (deviation drops below zero). Thus, the social attraction decreases. During the interval $[80000, 155000]$ the social attraction is reduced to zero because the relation between u and v was not active during this time interval. In the

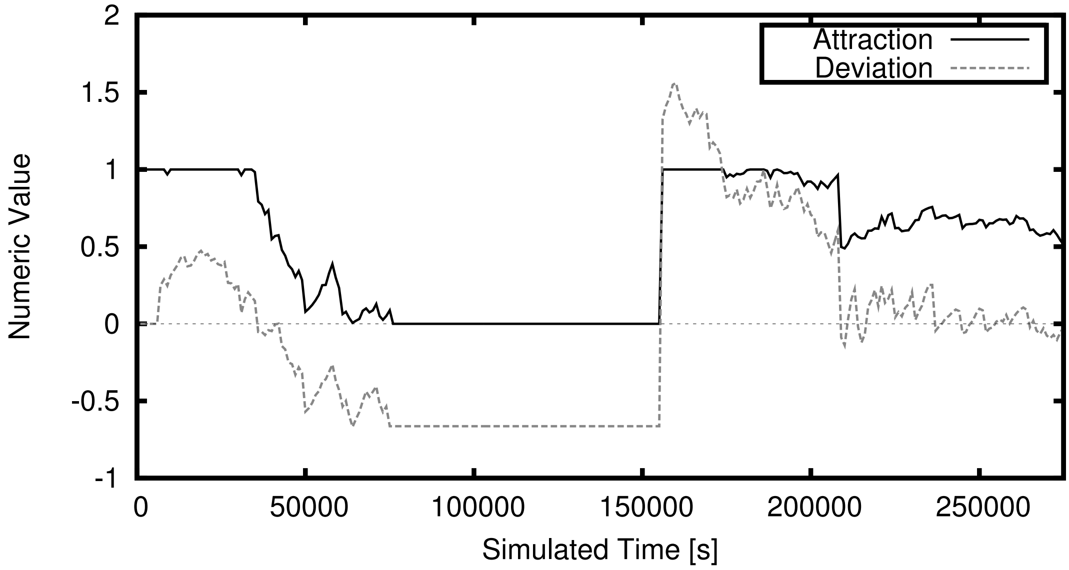


Figure 5.12: Example for the time-dependent behavior of the social attraction.

meantime, u and v produced many meetings with other nodes. Thus, afterwards, we observe a positive deviation peak because the number of produced meetings between u and v were too few compared to the number of meetings with other social acquaintances. Hence, the social attraction stays close to 1 which reduces the deviation and eventually leads to a good reflection (deviation drops down to zero) at $t = 205000$. Afterwards, the social attraction stabilizes at a value around 0.75.

5.6 Integration of Geographical Mobility Models

In the following, we identify several integration points for geographical mobility models and discuss the parameters that can be adjusted to match specific target scenarios.

In our prototypical implementation, nodes move directly from one anchor to the next in a straight line at constant velocity. At this point, one could integrate concepts of a geographical mobility model. For example, obstacles could be placed (e.g. buildings, walls, ...) between anchor points, which force nodes to move around them. Nodes may also be constrained by streets.

Because a constant speed among nodes may not be realistic in many scenarios, speed and acceleration models could be integrated. Note, however, that such models are highly dependent on the target scenario. For example, it was discovered that the speed distribution in campus environments follows a log-normal distribution [KKK06]. On the other hand, in a city simulation, nodes may move by car using other speed/acceleration models than pedestrians. Some nodes may be confined to streets, others to walkways.

Our model defines probabilistic group movements between anchors. However, we do not specify any micro-mobility behavior during the group movement. At this point, group mobility models could be integrated. For example, a reference point could be employed which moves from the source to the destination anchor. Nodes may move around this reference point randomly like in the reference point group mobility model (cf. Section 2.4.2).

Based on the current implementation of our model, nodes select a random goal within the geometrical boundaries of an anchor and move to this point. They stay there until their dwell time expires. This dwelling at anchors represents another possible integration point for a geographical mobility model. For example, some or all nodes could move within the anchor, based on some model (e.g. random walk). Or, one could integrate hierarchical models, similar to the two-level social mobility model (see Section 4.5). In such a model, *macro-anchors* may represent buildings. Within the geometrical boundaries of a macro-anchor, we could place micro-anchors which are visited by nodes currently dwelling at the corresponding macro-anchor. However, independent of the used geographical mobility model, it is important that nodes within the boundaries of the same anchor are in wireless range with significant probability or frequency. This is required to yield a proper reflection of the social network. For example, a random walk within the boundaries of the anchors would satisfy this requirement.

In the implementation of our proposed model, we place anchors randomly on the simulation area. However, the placement of anchors could be tailored to a specific target scenario. For example, if one would want to evaluate the use of a mobile application in a hospital environment, anchors could be created according to operating rooms, waiting rooms, or coffee machines. The geometrical boundaries of different anchors could be specified based on the locations they represent. Similar, anchor functions and dwell time distributions could be modeled with respect to the corresponding locations.

Even if no target scenario is known, empirical research could be used to create a probabilistic model for the placement of anchors. For example, it was discovered that at the macro-level (e.g. in a city), the length of human trajectories is well approximated by a truncated power-law distribution [GHB08]. Thus, one could implement an anchor placement algorithm which places anchors of the social sphere of a node such that their distances follow a truncated power-law distribution. This would essentially create an instance of our model that is coined for the simulation of city scenarios.

Some location may not be stationary, for example a bus. To model such 'locations', we could allow some anchors to move through the simulation area, e.g. based on regular schedules. In this case, the nodes dwelling at such an anchor would move relative to the anchor within its geographical boundaries.

In many scenarios, stationary infrastructure (e.g. access points) may be present in addition to nodes. If such a system model is used, our model could be adjusted to integrate a set of stationary nodes at specific locations, e.g. at anchors. For example, Chaintreau et al. defined a system model which consists of static and mobile nodes [CFL08]. They showed that social aspects might be utilized for efficient spatial gossip in such a system model.

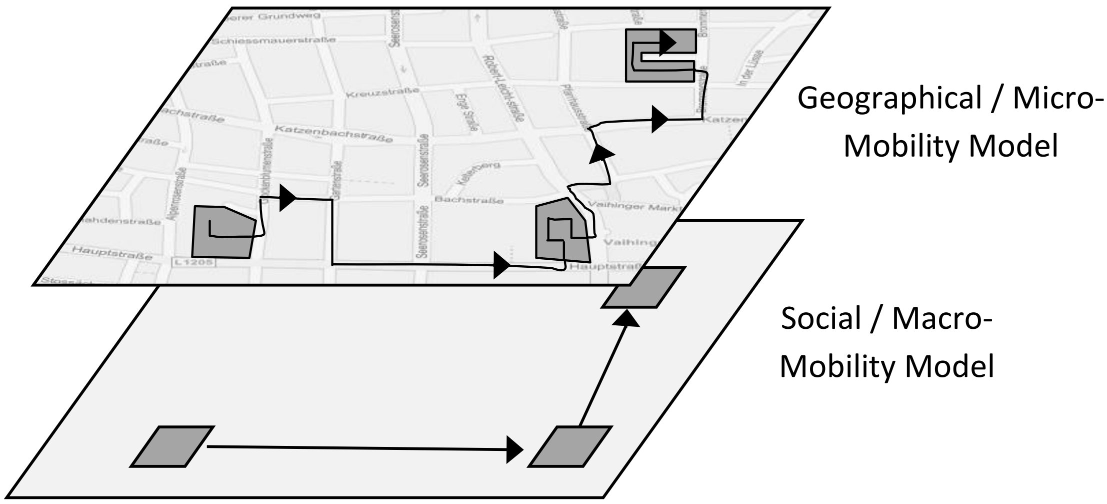


Figure 5.13: Integration of geographical aspects into the social mobility model.

We argue that the integration of concepts discussed in this section does not prevent the reflection of the social network. By employing the enhanced calculation of the node attraction proposed in Section 5.4.2, even if some nodes require much time between anchors caused by geographical concepts, they will still meet their social acquaintances. Based on concepts illustrated in the last section, some 'disturbances' in the reflection based on geographical concepts are balanced over time by adjusting the correction factor accordingly. We will verify this statement in the evaluation of our model.

Figure 5.13 visualizes the relation between the geographical and social mobility submodels. The social mobility model (lower plane) specifies that nodes move between anchors in order to reflect the input network. Thus, it defines the macro-mobility of nodes. The geographical mobility model (upper plane) on the other hand, focuses on micro-mobility. It specifies how abstract locations, i.e. anchors, are mapped to concrete locations. In this case, they are shaped according to real locations on a map. The geographical mobility model also maps the abstract movement between and within anchors to a geographical pathway. In our example, node movement is constrained to streets on a map and nodes move within the geographical boundaries of anchors according to some scheme (e.g. between rooms).

5.7 Generalized Social Mobility Models

As shown in the last section, we have designed our model to allow the integration of geographical concepts. In addition, we provide a large number of tunable parameters (summarized in Table 7.1 on page 81). By adapting these parameters and by integrating geographical concepts, our model may be specialized to yield the behavior described by other mobility models. We will demonstrate this ability in the following by describing how three existing social mobility models, discussed in Chapter 4, can be viewed as special cases of our model.

Note that in the following configuration descriptions, we assume that the simple calculation of the node attraction is employed. Since none of the generalized models features group movements, we set $p_{iso}(v,t) = 0 \; \forall v \in V, \forall t \in T$.

The parameters of our model can be adjusted to yield the behavior specified by the CMM (cf. Section 4.2). First, we have to generate social networks using the caveman model. By setting $\alpha = 1$ and $\phi = 0$ we let nodes move only according to the (simple) node attraction formula. To obtain a deterministic selection of the destination anchor, we have to set $g = \infty$. Thus, nodes always move to the anchor that exhibits the highest social attraction, which we calculate using the same way as in the CMM (cf. Equation 5.3). Initially, the positions of home anchors have to be randomly generated under the constraint that all home anchors of members of the same community share the same coordinates on the simulation area. Thus, all members of a community are initially grouped together like in the CMM. The dwell time of all anchors has to be set to zero since the CMM features no pause times.

To specialize our model to yield the HCMM (cf. Section 4.3), home anchors have to be placed as described above. In the HCMM, the probability that a node v choses the cell C as the next destination is proportional to the average weight between v and nodes having their home at C. This can be emulated by using $\alpha = 0$, $\phi = 0$, $g = 1$, and the same constant anchor function for each anchor. Because the home anchors of a community c are all placed on the same location on the simulation area, the probability to visit this location is proportional to the average weight of the social relation with nodes at c.

Finally, our model is also able to emulate the behavior of the two-level social mobility model (cf. Section 4.5). In this case, the social network consists of completely disjoint fully connected subgraphs which represent the communities. The dwell time distribution has to be set to a normal distribution. As defined by the two-level social mobility model, nodes must select their target anchor based on a Zipf distribution, which is shared among community members. Thus, we must set $\alpha = 0$, $g = 1$, and assign each anchor the same constant anchor attraction. The weights of the relationships with other members have to be selected from the corresponding Zipf distribution. Thus, the probability to move to a certain home anchor is heavy tailed as well. If all members of a community share the same attraction towards the same social acquaintances, they all share the same probability to visit certain anchors. Note that home anchors of members of different communities must be placed at the same location if both nodes have a non-zero probability to visit that location. In the previous section, we have already sketched how to integrate a geographical mobility model that adds aggregation points (i.e. a second level).

5.8 Summary

In this chapter, we have introduced our proposal for a social mobility model, which satisfies the requirements we have identified in Chapter 3.

The primary elements of the model are anchors, which represent abstract locations, and nodes, which move between anchors and stay there for a specific dwell time. To define the reflection of an input social network in a formal way, we have proposed the reflection requirement. The model produces meetings between nodes with a probability that is proportional to the weight of their social relations. The movement of nodes is driven by the attraction towards locations and nodes. However, nodes are attracted only to a subset of all anchors, which creates a spatial regularity. The location attraction exhibits a time-dependent characteristic, which creates temporal regularities. The node attraction towards nodes which stay at an anchor results in frequent meetings with social acquaintances. In some cases, nodes are repulsed by other each other to avoid meetings between nodes that would be based on the shared attraction towards common social acquaintances. Nodes do not always move alone from one anchor to the next. Instead, nodes probabilistically join other nodes to move as a group to a new destination anchor.

To improve the reflection of the input social network further, we identified several reasons why some social relations may be poorly reflected. We then have proposed concepts that extend the basic model described so far to cope with these problems. First, some structural properties of the input social network as well as geographical concepts may lead to a poor reflection of some social relations. The idea behind our solution is to employ a feedback control loop: We compare the number of meetings between nodes and adjust their social attraction according to the reflection requirement. This leads to a convergence towards a perfect reflection, even if some 'disturbances' are created, i.e. due to nodes traveling with different speed. Second, in a typical social network, the node degree is very heterogeneous among nodes (e.g. scale-free). Some nodes may only need to meet a few social acquaintances while hubs must produce many more meetings. This leads to a poor reflection of the social relation between pairs of nodes characterized by a very different node degree. To deal with this, nodes enter probabilistically into isolation phases. During an isolation phase, a node does not meet with other nodes for a specified duration. In general, the probability to enter an isolation phase is increased for nodes with only few social relations. This leads to a balance among the number of meetings per social relation and thus enables an optimal reflection of the social network. Algorithm 5.1 on the next page presents this extended model in pseudocode.

We have also sketched how geographical mobility models may be integrated into our social mobility model and reasoned, why this does not prevent the reflection of the social network. Finally, we have shown how to specialize our model to yield several existing social mobility models.

Algorithm 5.1 The Social Mobility Simulation Algorithm

```
procedure RunSimulation()
    RandomPlaceAnchors()
    V ← InitializeNodes()                    ▷ Place each node at one of its home anchors
    for t = 0, 1, .., t_max do
        if t ∈ {t_update, 2t_update, 3t_update, ...} then
            CalculateMeetingQuota()
        end if
        for all v ∈ V do
            if ReachedGoal(v) then
                        ▷ Count meetings with other nodes for the meet-quota calculation
                ReportMeetings(v)
            end if
            if LocatedAtGoal(v) ∧ DwellDuration(v) = 0 then        ▷ Choose a new goal
                if Random() < p_iso(v, t) then                ▷ Probabilistic isolation phase
                    EnterIsolationPhase(v)
                    DwellDuration(v) ← GetIsolationDwellDuration()
                else
                    G ← DetermineGroupMembers(v)
                                            ▷ Probabilistic Goal Selection
                    Anchor(G) ← SelectNewDestination(G)
                    for all v_g ∈ G do
                        Anchor(v_g) ← Anchor(G)
                        DwellDuration(v_g) ← SelectDwellDuration(Anchor(G))
                    end for
                end if
            else if LocatedAtGoal(v) ∧ DwellDuration(v) > 0 then
                DwellDuration(v) ← DwellDuration(v) − 1
                        ▷ Integration point for geographical mobility models
            else
                        ▷ Move towards goal according to the geographical mobility model
                MoveTowardsGoal(v, Goal(v))
            end if
        end for
    end for
end procedure
```

Chapter 6

Implementation

We have developed a prototypical implementation of our proposed model in C++. The implementation generates mobility traces which may be used by the NS2 simulator [NS2]. The implementation is originally based on the C implementation of the community-based mobility model [Uni] by Mirco Musolesi. We have modified and extended this implementation significantly to serve our needs. Our simulator may run either in a Windows or in a Linux environment.

In this section, we first show the generated trace format and illustrate how this trace integrates with NS2 (Section 6.1). Afterwards, we provide some details about the implementation of several selected concepts (Section 6.2). To conclude this chapter, we show an evaluation of the performance of the implementation (Section 6.2).

6.1 Trace Format and Integration with NS2

NS2 is a discrete event network simulator, which supports, among others, mobile adhoc networks. NS2 takes an OTcl script as input. OTcl is a simple object oriented script language. Such a script defines the topology of the network and the protocol/applications one wishes to simulate. The script is then used by NS2 to generate different kinds of output such as transmission delay, traffic, or lost packets.

In this thesis, we focus on creating a very dynamic topology. This may be realized by explicitly telling NS2 at which time two nodes may be connected, i.e. to simulate the connectivity graph. Another approach is to describe the movement of mobile nodes over time on a two dimensional rectangular area, using a specified transmission range. Based on this description, NS2 may simulate the emerging connectivity between nodes. In addition, this approach allows the integration of more sophisticated concepts like radio propagation models. Our implementation generates an OTcl script that describes the movement of mobile nodes. Listing 6.1 shows an example of such a generated output script.

Listing 6.1 Example for an OTcl movement trace, generated by our implementation.

```
$node_(0) set X_ 300.0
$node_(0) set Y_ 230.0
$node_(0) set Z_ 0.000000
$node_(1) set X_ 400.0
$node_(1) set Y_ 350.0
$node_(1) set Z_ 0.000000
...
$ns_ at 2.000000 "$node_(0) setdest 412.0 333.0 7.0"
$ns_ at 11.000000 "$node_(9) setdest 266.0 240.0 7.0"
...
```

First, the initial coordinates for each node are specified. For example, node 0 is placed at the coordinates $(300, 230)$ on the simulation area. The remaining lines in the script define the movement of the nodes. For example, node 0 starts to move towards the destination $(412, 333)$ at time $t = 2$ with a speed of $7m/s$.

This movement-trace may then again be imported by an OTcl script which defines the protocol/application layer based on the provided topology. Such a script is sketched in Listing 6.2. After setting some options, an instance of the NS2 simulator is created as well as the set of nodes. Afterwards, the simulation area is defined and the generated mobility trace is imported. Finally, a TCP connection between nodes 0 and 1 is simulated.

In addition to a NS2-specific output, our implementation is able to generate a general XML-based meta-format, provided by the original implementation of the CMM. This trace may then be parsed and transformed into arbitrary formats, for example by employing XSLT. In Listing 6.3, we show the same trace as described above using the corresponding XML output.

6.2 Implementational Details

We represented the input social network as an adjacency matrix, i.e. a two dimensional array. This network either may be provided as a plain text file or generated using several social network models. We chose this data structure because it allows for an efficient access to the weights of the individual relations. Note that the use of an adjacency *list* as representation for the social network may require less memory. However, based on the complexity of our model and the corresponding number of simulated nodes (< 150), we do not consider the memory requirements to be significant compared to the created performance improvement.

Listing 6.2 Example for an OTcl script, provided as input for NS2.

```
#set options
set val(chan)       Channel/WirelessChannel
set val(prop)       Propagation/TwoRayGround
set val(netif)      Phy/WirelessPhy
set val(mac)        Mac/802_11
set val(nn)         80  #num. nodes
...
set ns_    [new Simulator] #Creates an instance of the simulator

for {set i 0} {$i < $val(n) } {incr i} {
  set node_($i) [$ns_ node] #create nodes
}

set topo [new Topography] # setup topography object
$topo load_flatgrid 1000 1000 #create simulation area
...
#Load movement trace
source $val("movement_trace.tr")
...
# TCP connections between node_(0) and node_(1)
set tcp [new Agent/TCP]
$tcp set class_ 2
set sink [new Agent/TCPSink]
$ns_ attach-agent $node_(0) $tcp
$ns_ attach-agent $node_(1) $sink
$ns_ connect $tcp $sink
set ftp [new Application/FTP]
$ftp attach-agent $tcp
$ns_ at 10.0 "$ftp start"
...
```

Each anchor function is represented by a list of time/value pairs. These pairs are ordered according to an increasing time value. We only specify the anchor function for a single simulation period, i.e. all time values are in $\{0, 1, \ldots, t_p - 1\}$. If the anchor function needs to be evaluated at time $t \in T$ we first calculate $t' = t \pmod{t_p}$. Now, the list is traversed until a pair $(t_1, f_a(t_1))$ with successor $(t_2, f_a(t_2))$ is found such that $t_1 \leq t' < t_2$. Then, $f_a(t) = f_a(t')$ is calculated as the linear interpolation between $f_a(t_1)$ and $f_a(t_2)$, i.e.

$$f_a(t') = f_a(t_1) + \frac{f_a(t_2) - f_a(t_1)}{t_2 - t_1}(t' - t_1).$$

Listing 6.3 Example for an XML movement trace, generated by our implementation.

```
<?xml version="1.0"?>
<simulation xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<parameter>
   <xsize>1000.0</xsize>
   <ysize>1000.0</ysize>
   <numberOfNodes>80</numberOfNodes>
   <wirelessRange>20.0</wirelessRange>
</parameter>
<node_settings>
   <node>
      <node_id>0</node_id>
      <position>
         <xpos>300.0</xpos>
         <ypos>230.0</ypos>
      </position>
   </node>
   <node>...</node>
   ...
</node_settings>
<mobility>
   <position_change>
      <node_id>0</node_id>
      <start_time>2</start_time>
      <destination>
         <xpos>412.0</xpos>
         <ypos>333.0</ypos>
      </destination>
      <velocity>7.0</velocity>
   </position_change>
   <position_change>...</position_change>
   ...
</mobility>
```

This presents an easy and efficient way to implement the concept of a periodic anchor function. In general, more sophisticated methods are imaginable. For example, one could use a polynomial interpolation or splines. However, this may lead to performance degradations. Our implementation allows users to add further anchor function models adding only a few lines of code.

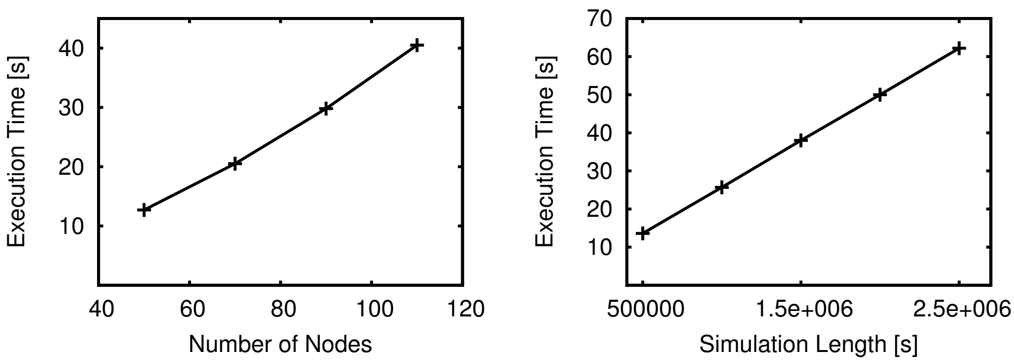
We have implemented several simple anchor function models, in particular:

- A sawtooth wave with parameterized height and number of spikes within the simulation period.
- A single spike within the simulation period (shown in Figure 5.4 on page 47) with parameterized height and random generated position.
- A constant function with a parameterized value. The use of this function for all anchors results in the absence of repetitive behavior.
- A plateau-like function characterized by a random generated position and parameterized height/plateau-length (shown in Figure 5.4 on page 47).

We implemented the concept of active social relations in a similar way. Each relation is represented by a function with domain $\{0, 1, \ldots, t_p - 1\}$ and image $\{0, 1\}$, where 0 means that the relation is not active and 1 means that the relation is active. We record the start and end intervals in which each relation is active within a single simulation period represented by time/value pairs in a list. To evaluate if a relation is active at time $t \in T$, we use the value $a \in \{0, 1\}$, provided by the last pair (t_1, a) in the list with $t_1 < t'$ and $t' = t \pmod{t_p}$. For example, the list $[(0,0), (300,1), (500,0)]$ assuming $t_p = 1000$ describes that the social relation is active in the time intervals $[300, 500]$, $[1300, 1500]$, $[2300, 2500]$, etc.

We implemented the enhanced calculation of the node attraction as introduced in Section 5.4.2 to evaluate the integration of geographical mobility models.

6.3 Performance of the Implementation



(a) ***Number of Nodes.*** (b) ***Simulation Length.***

Figure 6.1: Implementation execution time for different parameters.

We evaluated the performance of the implementation by measuring the execution time for a single simulation run. We conducted the evaluation on a Intel Core 2 Duo P8400

(2,26Ghz) CPU. Note, however, that our implementation does not utilize parallel threads. Thus, there may be room for further performance improvements. Each point in the following plots represents the average over 100 measurements. Figure 6.1(a) shows the impact of the number of nodes (y-axis) against the execution time (x-axis) by using a simulation length of $t_{max} = 10^6$. Note that in this case, the order of magnitude of the total number of contacts between nodes is about 300000 for $n = 80$ nodes. By increasing the number of nodes, we observe only a very low polynomial increase of the execution time. This shows that an efficient implementation of our model is possible despite its complexity.

Figure 6.1(b) shows the execution time against the simulation length for $n = 80$ nodes. As expected, we observe a linear dependency. This shows that even very long movement traces can be generated with acceptable execution time. Note that the execution time of our mobility trace generator is significantly smaller compared to the time NS2 requires to simulate the generated trace subsequently.

Chapter 7

Evaluation

This chapter presents the evaluation of our implementation. First, we discuss the methodology. Then, we verify the requirements we have identified in Chapter 3 and explore the parameter space of our model. Finally, we give a summary of the results and their implications.

7.1 Methodology

The goal of this evaluation is to show that the identified requirements (see Chapter 3) for a realistic social mobility model are satisfied. To accomplish this, we use a metric to measure each requirement. We use existing metrics if possible, to compare our results to characteristics of real traces.

7.1.1 Simulation Setup

For each node, we created a single home anchor. All anchors are placed randomly on an equidistant grid. We use the *unit disc graph model* [CCJ90] to determine connectivity. That is, we assume that two nodes are in contact if their Euclidean distance is below a fixed transmission range. The distance between anchors on the grid is chosen in a way such that nodes at neighboring anchors are not in transmission range, but nodes within the boundaries of the same anchor are. This standard anchor layout is shown in Figure 7.1.

We used a uniform distribution in the interval $[200, 300]$ seconds as the dwell time distribution for each anchor. However, we also performed experiments using a power-law distribution and observed very similar qualititative results.

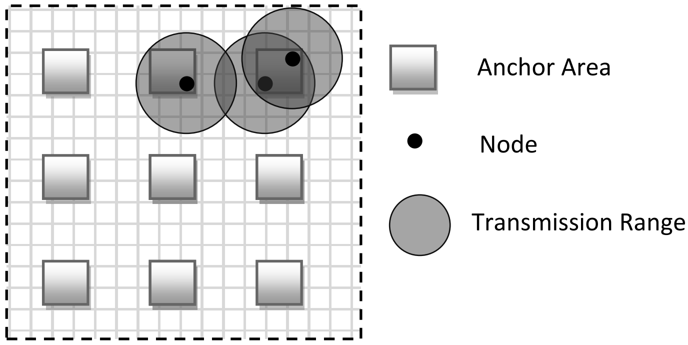


Figure 7.1: Simulation layout used for the evaluation of the model.

Furthermore, we employed a plateau-function, similar to the grey curve shown in Figure 5.4 on page 47 to describe the anchor functions. In particular, we used the following anchor function for each $a \in A$:

$$f_a(t) = \begin{cases} 0.8, \text{ if } t_0 \leq t \leq t_0 + t_c \pmod{t_p} \\ 0.01, \text{ else.} \end{cases}$$

We used a fixed plateau-length $t_c = t_p/4$ and set the position of the plateau for each anchor to a point chosen uniform randomly from $t_0 \in \{0, 1, \ldots, t_p - 1\}$. Note that we have obtained very similar results for other implemented anchor functions (cf. Section 6.2), except for the constant anchor function which does not create temporal regularities. However, this does not necessarily mean that the anchor function is not relevant. It only means that the realistic characteristics of our model are independent of the anchor function. To simulate specific target scenarios realistically, it is worthwhile to shape the latter appropriately.

Over the course of many experiments during the implementation phase, we have found a set of reasonable good standard parameters. Table 7.1 shows the standard simulation setup. In this table and in the remaining chapter, the unit $[s]$ denotes a single simulation tick. The parameters were selected such that a simulation tick roughly corresponds to a second in the real world, which enables a comparison to real mobility traces. We used these parameters for all conducted simulations, unless explicitly stated otherwise.

We also explored the parameter space by analyzing isolated parameters while keeping all other parameters fixed at the standard value. However, because of space related reasons, we only show the important results in this evaluation.

A primary goal of this evaluation is to show that the validity of the results is independent of the input network. To accomplish this, we evaluated the model using several different social network models as input. The parameters of the individual models were chosen such that the average node degree is about the same for all models if the standard parameters are

Parameter	Standard Value	Page Reference
Simulation Area Size	$1000m \times 1000m$	39
Number of Nodes	$n = 80$	5
Node Velocity	$7m/s$	39
Transmission Radius	$20m$	79
Simulation Duration	$t_{max} = 2000000s$	43
Social Network Model	Toivonen model	11
Node/Location Attraction Weight	$\alpha = 0.5$	54
Greediness	$g = 1$	54
Penalty Factor	$\phi = 0.3$	54
Penalty Threshold	$w_{weak} = 0.5$	53
Group Movement Threshold	$g_{thresh} = 0.7$	45
Simulation Period	$t_p = 100000s$	43
Length of Relation Active	$t_p/2$	82
Meeting Quota Update Interval	$t_{update} = 1000s$	62
Isolation Phase Length	$250s$	58

Table 7.1: Standard parameter setup for the evaluation of the model.

used. If this would not be the case, it would be difficult to assess whether varying results are due to the different node degree or due to the different structure. In particular, we used the following models (cf. Section 2.2.4):

- **The Caveman Model.** Caves of 16 nodes were created, independent of the number of nodes, which leads to a (constant) node degree of 15. We used a rewiring parameter of 0.2. This yields a very strong clustering coefficient. As we will show in this chapter, the clustering coefficient exerts a significant influence on some of the used metrics.

- **The Holme-Kim (HK) Model.** Initially, a random network of 20 nodes was created using a probability of 0.5 to create an edge for each pair of nodes. In every subsequent step, 11 edges to existing nodes were created with a triad formation step probability of $p_t = 0.25$.

- **The Toivonen Model.** Initially, a random network of 20 nodes was created as for the HK Model. In each step, the number of initial random attachments was set to $m_r = 2$ with a probability of 0.9 and to $m_r = 3$ with a probability of 0.1. The number of secondary contacts was selected uniform randomly from $m_e \in \{5, 6, 7, 8, 9\}$. This corresponds to parameters for which Toivonen et al. yielded good results.

The properties of the generated social networks are shown in Table 7.2. We measured the average value over 1000 created instances. Lengths of the 95% confidence intervals are enclosed in brackets. Note that we have performed experiments with further network models (e.g. the random graph model and the BA model) and observed very similar characteristics. However, to maintain a clear depiction, we have only shown the three models above because they capture different typical properties of social networks.

Social Network Model	**Node Degree**	**Clustering Coefficient**
Caveman	15.0(0)	0.543(0.109)
Toivonen	15.576(1.200)	0.364(0.051)
Holme-Kim	15.402(0.625)	0.295(0.038)

Table 7.2: Properties of the generated social networks.

Currently, no social network model exists that models the strength of the social relationships or time intervals in which a social relation is active. Thus, we assigned each relation a uniform random weight from the interval $]0, 1]$. We describe the assignment of time intervals in which a social relation is active in the following.

To satisfy the characteristic of human habitual behavior, social relations should be active periodically. For example, a relation between two colleagues is typically active every Monday during working hours. In our model, the total simulation duration is partitioned into simulation periods of length t_p. The idea behind the generation of $\chi(.)$ is to assign each social relation a uniform random time interval of constant length *within the simulation period*.

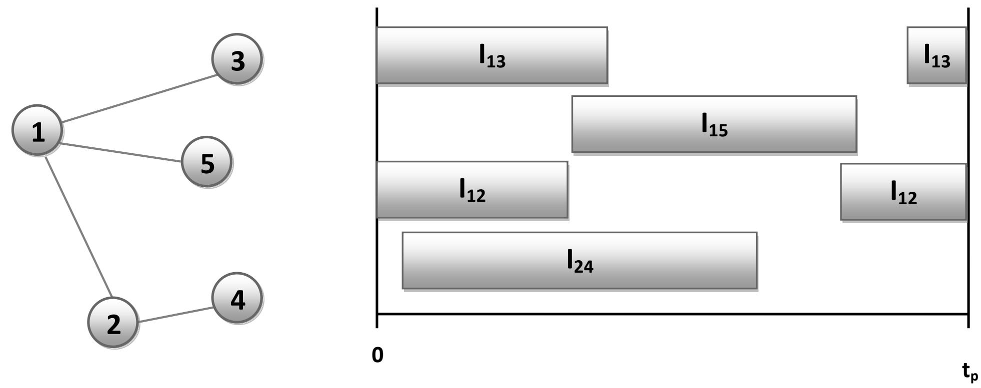


Figure 7.2: Example assignment of active social relation intervals within a single simulation period.

An example of such random assignment is shown in Figure 7.2 for a single simulation period and four relations in an example social network (left). The intervals within one simulation period in which the relation between two nodes v_i and v_j is active (right) is denoted as

$I_{ij} \subseteq [0, t_p]$. After the assignment of the time intervals in which a relation is active within *one* simulation period, the time intervals $\chi(.)$ within the whole simulation duration for two nodes v_i, v_j is defined as follows:

$$\chi(v_i, v_j) = \{t \in T \mid t \pmod{t_p} \in I_{ij}\} .$$

In other words, a social relation at an arbitrary time $t \in T$ is active if the social relation is active at the time within the current simulation period $t \pmod{t_p}$ according to the discussed random generation. This yields periodically recurring intervals.

7.1.2 Measurement of the Simulated Mobility

Contacts and anchor visits were measured during the simulations to provide an input for the metrics used in this chapter. For each stay at an anchor, we recorded the start/end timestamp and the ID of the corresponding anchor. In addition, we measured contacts between nodes. Note that in contrast to *meetings* between nodes, which refers to overlapping dwell time intervals at anchors, contacts may be created on the way between anchors by random encounters. We measured contacts between nodes as follows: At periodic time intervals, the positions of all nodes are compared to each other. If two nodes are within transmission range (as explained above) and the nodes are not already in a contact state, the current time is recorded and both nodes enter into the contact state with each other. If two nodes are in a contact state and if they are not in transmission range anymore, the current time is recorded and both nodes leave the contact state with each other again. We chose this method because it is very similar to the methodology used to measure direct contact traces (see Section 2.4.1). In particular, this allows the comparison of the inter-contact distribution to such mobility traces. In addition, this method is quite efficient which allowed us to execute many simulations to yield small confidence intervals.

Each point in the plots and each distribution presented in this chapter is based on 100 simulation runs, if not stated otherwise. We show the mean value and 95% confidence intervals (shown as error bars). The distributions yielded too small confidence intervals to be clearly depicted. In the following, all shown distributions have a mean 95% confidence interval of ≤ 0.001, if not explicitly stated otherwise

7.2 Reflection of the Input Social Network

The goal of this section is to validate the reflection of the input network. Furthermore, we explore the influence of different concepts like node repulsion and the correction factor on the reflection. We also simulate the effects of geographical mobility models and their impact on the reflection of the social network. However, we have to propose an appropriate metric first.

Social Network Model	**Fraction of Pairs Matched**	95% Confidence Interval
Caveman	0.9991	0.0015
Toivonen	0.9989	0.0011
Holme-Kim	0.9990	0.0011

Table 7.3: Results of the reflection metric for different social network models.

7.2.1 Reflection Metric

Our metric is based on the following idea: Assume that a node u has social relations to nodes v and w. Furthermore, assume that the weight of the social relation with v is greater than the weight of the social relation with u. Based on this assumption, we would expect that u has created more contacts with v than with w during a simulation of significant length. Furthermore, we would expect that any pair of social acquaintances produces more contacts than any pair of nodes without a social relation. Note that these assumptions correspond to our definition of the reflection requirement (cf. Definition 6 on page 43). The idea behind the design of the metric is to validate these assumptions for all such *node triples*.

Let $c(u, v)$ denote the number of measured contacts between u and v during a complete simulation run. For all node triples

$$N = \{(u, v, x) \mid \{u, v\}, \{u, x\} \in E : w(u, v) > w(u, x) + \mu\}$$

with $\mu \in [0, 1]$, we verify (after each simulation run) if

$$(u, v, x) \in N \Rightarrow c(u, v) > c(u, x) \tag{7.1}$$

holds. Note that Statement (7.1) is only evaluated for two social relations whose weight has at least a difference μ. If two social relations have nearly the same weight, the statement may be violated without necessarily representing a poor result. For example, if $w(u, v) = 0.5$ and $w(u, x) = 0.51$, a result of $c(u, v) = 101 > c(u, x) = 100$ represents a rather good reflection because two social relations of nearly equal weight produced nearly the same number of contacts. To avoid obscuring our results, we do not consider pairs of relations with a similar weight. We used $\mu = 0.1$ for all our simulations.

The result of the so called **Reflection Metric** $R(N)$, is the *fraction* of all node triples in N, for which Statement (7.1) holds.

7.2.2 Initial Results and Refinement of the Metric

Table 7.3 shows the reflection metric for our standard simulation setup using different social network models to generate input networks. Note that a mobility model based on purely

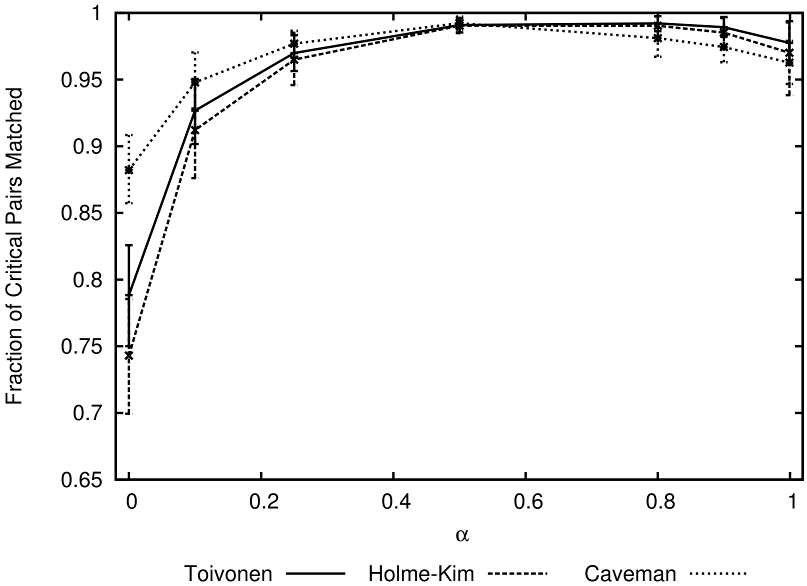


Figure 7.3: Impact of α on the Reflection Metric

random encounters (e.g. random walk/waypoint) yields a result of $R(N) = 0.5$ for $t \to \infty$. The results suggest a very good overall reflection. To show the effects of different parameters on the reflection metric more clearly, we additionally evaluated $R(N_{critical})$ with

$$N_{critical} = \{(u, v, x) \mid \{u, v\}, \{u, x\} \in E :$$
$$w(u, v) > w(u, x) + \mu \wedge |w(u, v) - w(u, x)| < 2\mu\} \subseteq N.$$

This corresponds to relations which have a certain difference ($> \mu$), yet are quite similar (difference $< 2\mu$). We call the set $N_{critical}$ the **critical relations**. The name relates to the fact that such pairs of social relations are difficult to reflect appropriately. Since they have similar weights, it is more probable that $c(u, v) > c(u, x)$ does not hold. Hence, this metric is more sensitive to parameter changes that degrade the reflection of the social network.

Figure 7.3 shows $R(N_{critical})$ (y-axis) for $\alpha \in [0, 0.999]$ (x-axis). The results are depicted for all three different models used to generate social networks. As discussed in Section 5.4.4, α weights between the node and location attraction. Note that $\alpha = 1$ is not a sensible option because then a node would never move to an anchor that has no associated nodes. This leads to the undesired behavior discussed for the CMM-Model (cf. Section 4.2).

The qualitative behavior is the same for all used social network models. We observe an increasing reflection if α increases until $\alpha = 0.5$, followed by a constant, nearly perfect reflection. If α is increased beyond a value of 0.8, we observe a slight decrease of the reflection. This shows that at least some amount of node and location attraction is required to yield an optimal reflection. In particular, for a large range of values, i.e. $\alpha \in [0.25, 0.999]$, we observe a match of over 95% for $R(N_{critical})$ and over 99.5% for $R(N)$ (not shown in the plot).

Furthermore, we observe that for $\alpha < 0.25$ the caveman model yields the best results followed by the Toivonen-model and the HK model, respectively. We suspect that the reason for this lies in the structural properties of the input network. For small α, nodes primarily move between home anchors of social acquaintances irrespective if they might actually meet them there. Based on the strong community structure and clustering of the caveman model, nodes seldomly meet nodes which are not social acquaintances at such anchors. However, such encounters are more probable if the HK or Toivonen model is used to generate social networks because they exhibit a less pronounced clustering and community structure.

The results shown so far have also shown that random encounters on the way between anchors do not seem to have a significant effect on the reflection of the social network. We have investigated such random encounters and discovered that their order of magnitude is comparable to the total number of meetings between nodes at anchors. However, they are equally distributed among all pairs of nodes (i.e. they are in fact random). Note that neither the reflection requirement nor the reflection metric make any assumption on the *total* number of contacts between nodes, i.e. both allow random encounters. We consider this to be realistic because people typically do not only meet their social acquaintances, but also have random encounters with other individuals. Such contacts may also be utilized by mobile applications. However, they are less frequent and typically of a shorter duration.

We have also performed simulations using different numbers of nodes, shown in Figure 7.4 for $R(N_{critical})$. Only a very weak linear degradation (below 0.003) is observable for the whole range of values using the HK or Toivonen model to generate social networks. The reflection of the caveman model even increases over time. Thus, we yield good results for a reasonable large number of nodes. This is an important result because the number of nodes reflects different mobility scenarios, and a mobility model should allow the variation of this parameter.

7.2.3 Node Repulsion

We introduced a repulsion between nodes to avoid that two nodes which do not share a strong social relation meet too frequently because they are both attracted to one or more common social acquaintances (cf. Section 5.4.3). The strength of this repulsion is controlled by the penalty factor ϕ. Figure 7.5 shows $R(N_{critical})$ for different values of ϕ. $\phi = 0$ corresponds to the case that nodes do not exert any repulsion towards each other.

We observe that the penalty factor significantly improves the reflection of the social network until about $\phi = 0.3$. After that, ϕ seems to have no significant effect. We also notice that the reflection of social networks generated from the caveman model seems to be less affected by a weak node repulsion. Again, this characteristic may be explained by to the strong

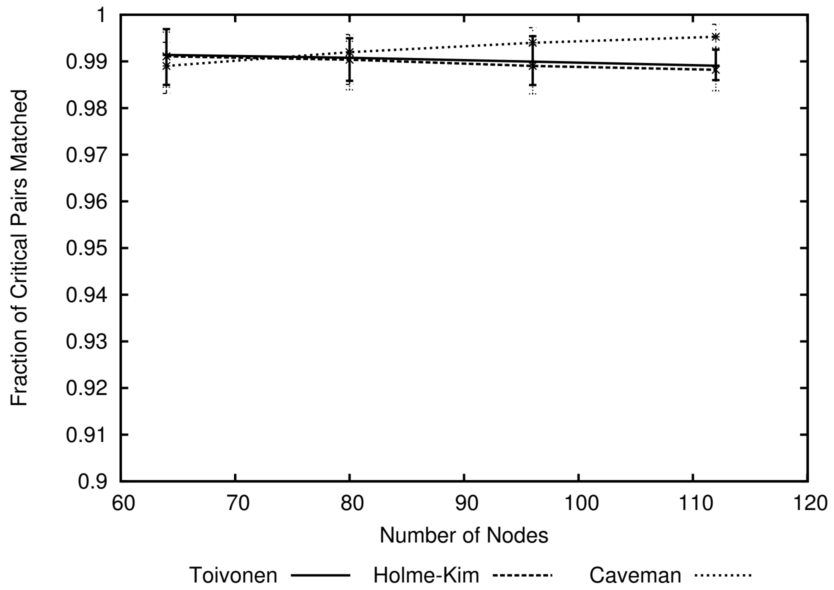


Figure 7.4: Social network reflection for an increasing number of nodes.

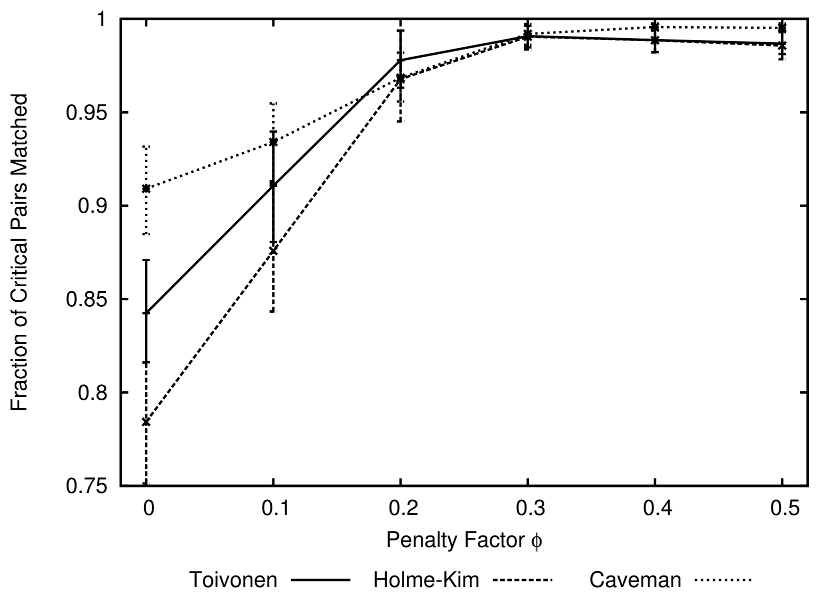


Figure 7.5: Social network reflection for an increasing ϕ.

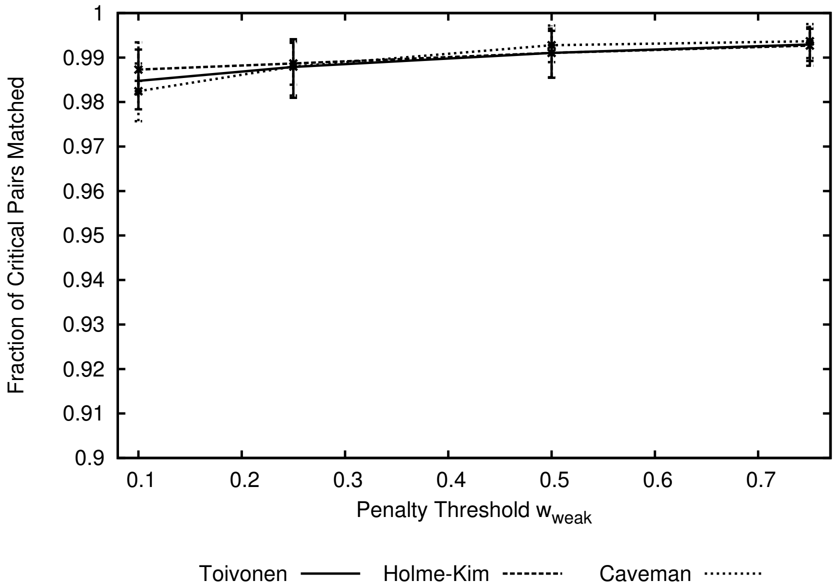


Figure 7.6: Reflection metric for an increasing penalty threshold w_{weak}.

clustering coefficient of the caveman model. For social networks generated by this model, the probability to find nodes that share a common social acquaintance but have no social relation of their own is less probable. Hence, it is more seldomly required that nodes are repulsed by each other to avoid undesired meetings. Note that for $\phi = 0.3$, the reflection seems to be quite well irrespective of the used social network model.

The second parameter associated with the node repulsion is the penalty threshold w_{weak}, which essentially gives us the definition of a 'weak' social relation. As discussed in Section 5.4.3, all social relations below this threshold are considered to be weak (including the case of no social relation) and thus may cause a node repulsion. Figure 7.6 shows the impact of the penalty threshold on $R(N_{critical})$. It can be observed that an increasing penalty threshold improves the reflection of the social network only slightly, compared to the improvement that the introduction of the node repulsion provides in the first place.

In the following, we try to explain this effect. Assume that a relation with a strength beyond a small threshold exists between two nodes u and v. Further assume that too many meetings between u and v were produced because at least one common social acquaintance between u and v exists. In this case, the social attraction between u and v is reduced. This leads to less meetings in the following simulation ticks and balances the number of meetings over time towards the appropriate level (as defined by the reflection requirement). However, this is not possible if two nodes do not have a social relation at all, because their ideal number of meetings is below the number of meetings of any two social acquaintances.

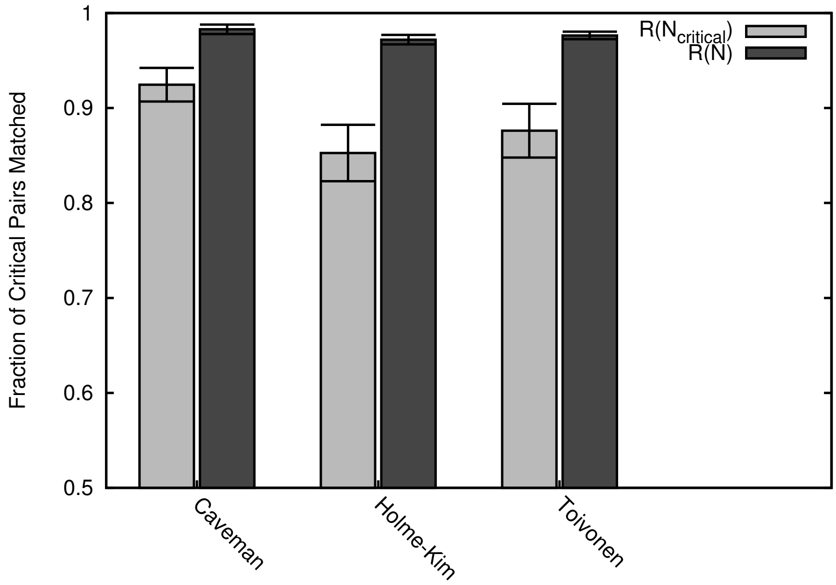


Figure 7.7: $R(N)$ and $R(N_{critical})$ without correction factor and isolation phases.

Essentially, the node repulsion does *avoid* meetings between nodes with a weak or non-existing social relation in certain situations. Because we do not want to prevent meetings which do not weaken the reflection of the social network, we consider a penalty factor of $\phi = 0.3$ as the optimal value.

7.2.4 Robust Reflection of the Social Network

In the following, we show that the extension of our model by incorporating the correction factor and isolation phases (cf. Section 5.5) yields a significantly better reflection of the social network. In Figure 7.7, we show $R(N)$ and $R(N_{critical})$ *for the basic model* using different social network models . A significantly decreased reflection for $R(N_{critical})$ is observable compared to the extended model (cf. Section 7.2.2). However, the overall reflection seems to be still acceptable. This shows that the problems related to the heterogeneous sociability and structural properties of the social network primarily affect critical relations.

We also observe that the caveman model seems to be the best reflected social network model, followed by the Toivonen and HK model, respectively. This can be explained due to the structure of the generated networks. The caveman model is characterized by a constant node degree, which makes the concept of isolation phases unnecessary. The node degree distribution of the Toivonen model, on the other hand, is characterized by a power-law followed by a cut-off. The HK model creates scale-free networks. Thus, isolation phases are

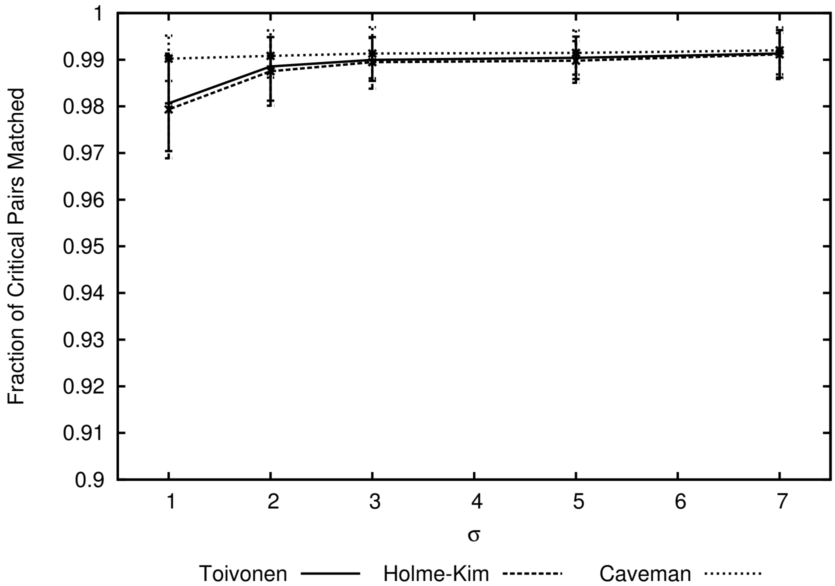


Figure 7.8: Reflection metric against σ.

necessary to prevent the degradation of the reflection. In addition, the caveman model is characterized by a simple structure based on distinct isolated communities and therefore by a very homogenous similarity among nodes. The other models, however, have a more complex structure which may lead to disruptions.

Note that if we use the extended version of our model (cf. Table 7.3), the reflection is quite well, irrespective of the used social network model. This shows that these concepts indeed fulfill their requirements to make the reflection more robust against structural idiosyncrasies.

In the following, we will show that our model also yields a good reflection in the presence of certain geographical concepts. We simulated this by assigning each node a fixed constant speed, uniformly selected from the interval $[\sigma, 7]m/s$ with $\sigma \in \{1, 2, 3, 5, 7\}$. This simulates the effect that some nodes require more time to travel between anchors due to geographical concepts, like obstacles or long distances. Figure 7.8 shows the impact of heterogeneous velocities among nodes on $R(N_{critical})$. Only a small degradation for a decreasing σ is observable. Note that even in the case of $\sigma = 1m/s$, where some nodes may require much more (by a factor of 7) time to travel between anchors, the simulation still yields an average match of over 98% for $R(N_{critical})$ and over 99.7% for $R(N)$.

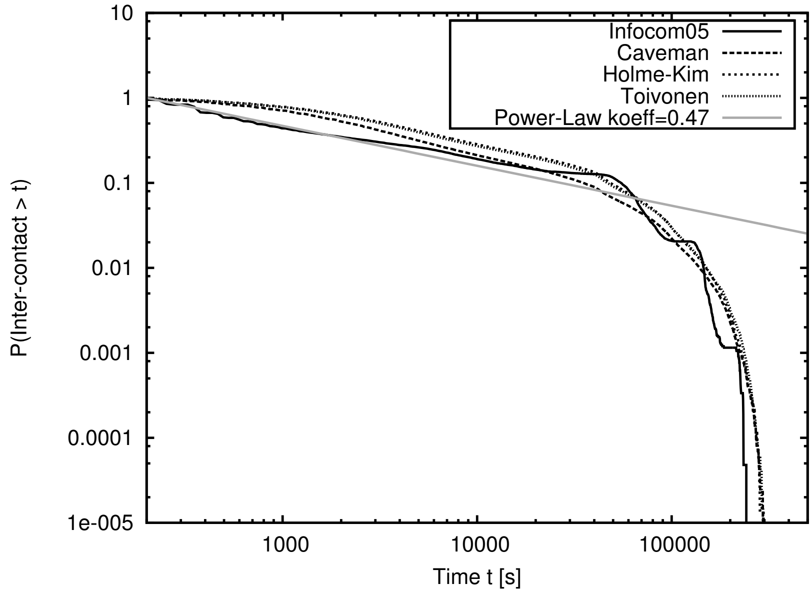


Figure 7.9: Inter-contact CCDF for different input social networks in comparison to a real trace.

7.3 Inter-contact Distribution

To validate the realism of our model, we compare the inter-contact distribution of our model to the inter-contact characteristic of real mobility traces in this section. The latter typically exhibit a power-law tail until about half a day. After that, an exponential degradation follows. The length of this exponential tail is approximately proportional to the length of the simulation. For comparison, we use a mobility trace acquired during the INFOCOM 2005 conference that we already discussed in Section 2.4.1. We chose this mobility trace since its characteristics are representative of the characteristic dichotomy observed in other mobility traces. In addition, most other real mobility traces are AP-based. We have found no other publicly available direct contact trace with a comparable number of contacts that features more nodes.

To adjust our model to the fact that the INFOCOM-trace was recorded for three days, we simulated only three simulation periods. The inter-contact CCDF of our implementation is shown in Figure 7.9 for the different social network models. We use a logarithmic scale for both axis to depict power-laws more clearly. It is observable that the inter-contact distribution of our implementation displays very similar characteristics compared to the inter-contact distribution of the INFOCOM-trace. In particular, a distinct power law until about $t \approx 50000$ is observable. For a better comparison, we also plotted a power law with a coefficient of $\lambda = 0.47$. This value falls into the typical range of real mobility traces, where

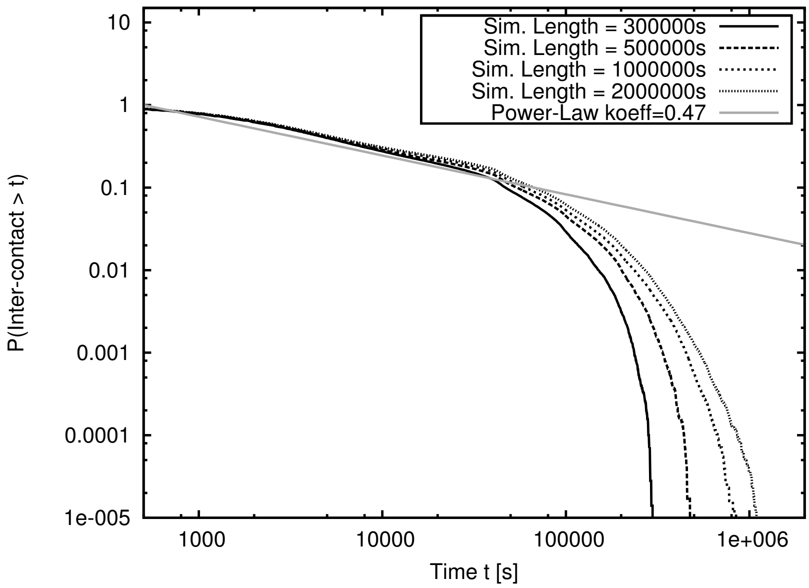


Figure 7.10: Inter-Contact CCDF for different simulation durations.

coefficients are found to be between 0.2 and 0.7. After a characteristic time of about half a day (or simulation period), we observe an approximately exponential degradation. This corresponds to the dichotomy found in real mobility traces

The length of the exponential tail in real mobility traces is strongly dependent on the length of the experiment. To verify the presence of this property in our model, we conducted simulations using different simulation durations. As shown in Figure 7.10, the length of the exponential tail increases as the length of the simulation increases, similar to real mobility traces.

The presence of a power-law seems to be an emergent property of our model. However, we have investigated which concepts are important to yield a heavy-tailed inter-contact distribution and identified two reasons.

First, we have found that the concept of active social relations is necessary. Without this concept, the inter-contact distribution follows a fast exponential decay. A requirement to create a heavy-tailed inter-contact distribution is to have a large number of long inter-contact times. Two nodes typically do not meet if their social relation is not active which creates the latter. Note that we modeled this concept to resemble real-world behavior of humans (cf. Section 3.2) and actually gained realism in terms of the inter-contact distribution.

Second, we have observed that a penalty factor of $\phi > 0$ is required. The actual value seems to have little importance to gain power-law characteristics. $\phi = 0$, on the other hand, results

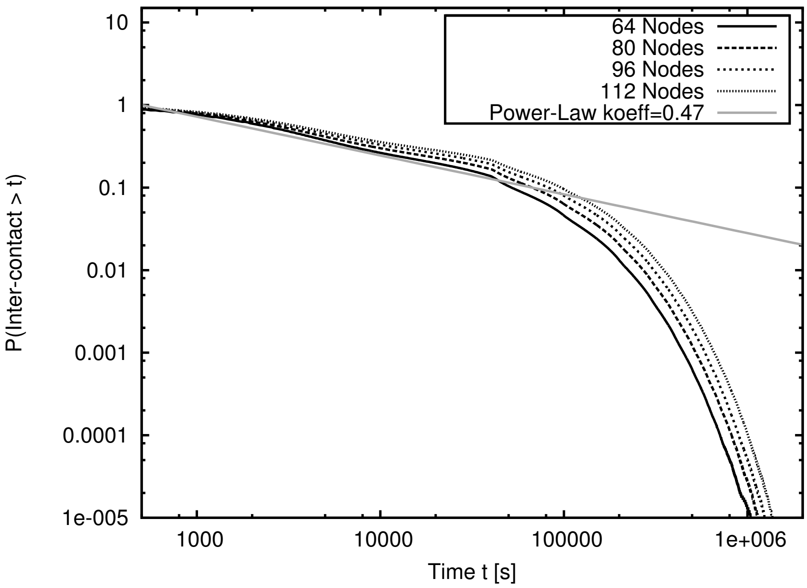


Figure 7.11: Inter-contact CCDF for different numbers of nodes.

in a behavior that resembles rather a long exponential tail. The reason for this is that long inter-contact times are frequently produced by pairs of nodes which have only a weak social relation or no social relation at all, because their probability to meet is low. However, in the absence of node repulsion, many such nodes meet frequently based on common social acquaintances and thus may not produce long inter-contact times.

We also evaluated the influence of the number of nodes on the inter-contact distribution. This is shown in Figure 7.11. We observe that the discussed characteristic of the inter-contact distribution is independent of the number of nodes. However, the power-law coefficient seems to decrease slightly if we increase the number of nodes. This is due to the fact that by increasing the number of nodes, we increase the number of pairs of nodes without a social relation. Such pairs typically produce long inter-contact times, which explains the slightly more heavy-tailed distribution. By comparing different real mobility traces, we observe that an increased number of nodes typically does also lead to a decreased power-law coefficient [CHD+07]. We suspect that this is due to the same reason.

Some authors validate their model by comparing their results to the contact duration distribution of real mobility traces. However, this distribution seems to be rather scenario-dependent and different characteristics are possible [HH05, KKK06]. The advantage of our model is that the contact duration distribution may be arbitrarily shaped by providing an appropriate dwell time distribution. If, for example, a power-law contact duration distribution is desired, the anchor dwell time distribution of all anchors could be set to such a power-law distribution (with a greater mean however).

7.4 Spatial Regularity

As discussed in Section 3.3, spatial regularity refers to the characteristic of humans to visit some locations regularly (i.e. the social sphere of a node) and most other locations only with low probability.

To evaluate the spatial regularity exhibited by our model, we created an *anchor preference distribution* as follows: First, we define the anchor preference of an anchor $a \in A$ with respect to a node $v \in V$ as the ratio between the number of visits of v at a and the sum of all visits v has performed (at arbitrary anchors). All anchors are ranked according to the anchor preference, where the anchor with the L-greatest anchor preference gets rank L. Finally, we calculate the average anchor preference over all nodes for each rank to yield an anchor preference distribution.

Because in our experiments the average dwell time was the same for all anchors and all nodes moved with the same constant speed, v's anchor preference for a may also be regarded as an approximation of the probability to find v at a, averaged over *the whole* simulation duration. We call this probability the **co-location probability**. Gonzales et al. have measured the co-location probability for a large data set (cf. Section 3.3) which allows us to make a comparison.

Figure 7.12 shows the anchor preference distribution using different social network models on a log-log scale. The increasing anchor rank (x-axis) against the anchor preference (y-axis) is depicted. It is observable that a node typically spends clearly the most time at one anchor – its home anchor. For the remaining ranks, we observe a heavy tailed distribution. This is very similar compared to real traces, which also exhibit a heavy tailed preference of locations (see Section 3.3). Independent of the used social network model, nodes seem to spend most of their time at a few anchors. For comparison, we have also drawn Gonzales et al.'s approximation $P(L) = c/L$ such that $\sum_{L=1}^{L=n} c/L = 1$ (solid grey). We observe that our model captures the spatial regularity of this empirical study quite well.

The HK and Toivonen social network models show very similar characteristics. If the caveman model is used to generate social networks, on the other hand, the distribution is slightly more heavy tailed. In particular, the probability to find a node at its home anchor (rank 1) is decreased while the probability to find a node at anchors of the ranks $2 - 19$ is increased compared to the other used models for social networks. However, for ranks > 20 we observe a faster decay. We have identified the following reason for this behavior: As discussed in Section 7.2.3, an increased clustering coefficient leads to a decreased probability that nodes are repulsed by other nodes. Thus, 'foreign' anchors are more attractive to nodes. In addition, it is less likely that nodes move to anchors which are not in their social sphere

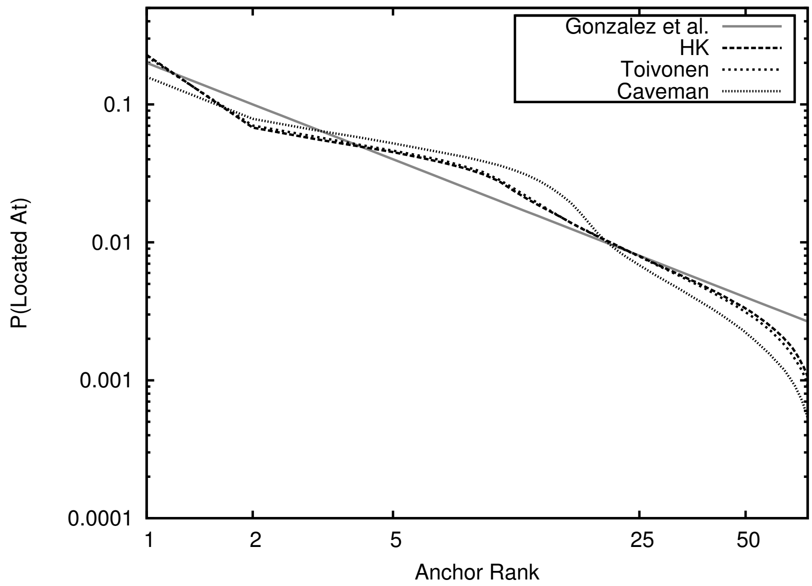


Figure 7.12: Anchor preference distribution using different models for social networks.

because the social spheres of social acquaintances overlap strongly. Thus, a node visits home anchors of social acquaintances (in our example typically ranks $2 - 16$) more frequently if the social network has an increased clustering coefficient.

α, which weights between the node and location attraction, seems to have a strong influence on the anchor preference distribution, as shown in Figure 7.13. Independent of α, we observe a heavy tailed distribution. However, decreasing α seems to lead to a steeper tail. We have identified the following reason: For $\alpha \rightarrow 0$, nodes are strongly influenced by their location attraction. Thus, they only move between home anchors of their social acquaintances with a probability proportional to the strength of their relationship. For $\alpha \rightarrow 1$ however, nodes are strongly influenced by their social attraction. If a social acquaintance u of node v is located at an anchor a, a will exert a strong attraction towards v, even if a is not a member of v's social sphere. Thus, increasing α increases the probability that nodes move to anchors which are not part of their social sphere, which leads to a more heavy tailed anchor preference distribution.

7.5 Temporal Regularity

An important requirement to gain realistic mobility characteristics and to enable the evaluation of predictive applications, is to simulate temporal regularities (cf. Section 3.4). To create a temporal regularity on a global level, we have introduced a periodic temporal

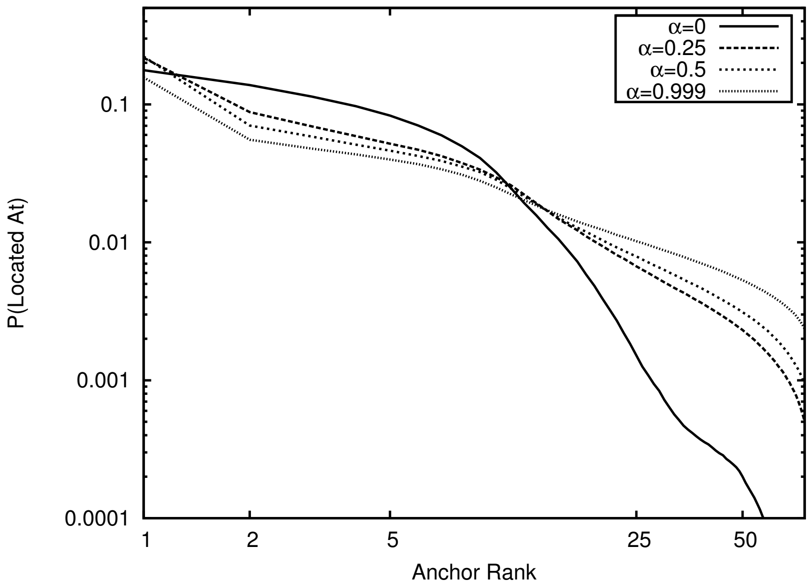


Figure 7.13: Anchor preference distribution against α.

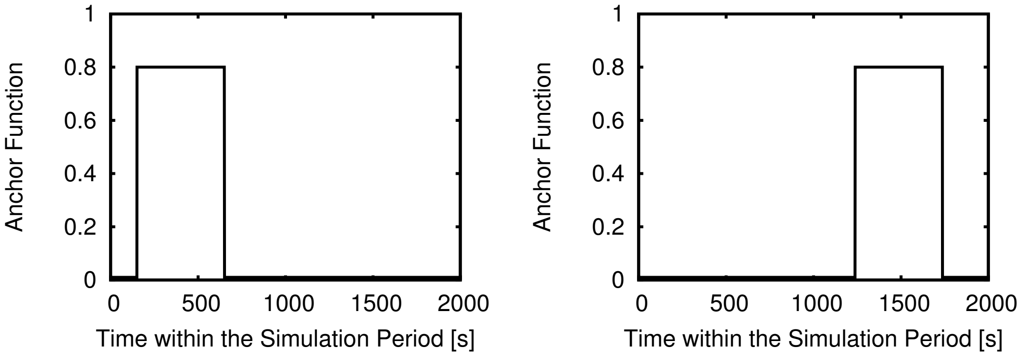
dependency by employing anchor functions (cf. Section 5.4.1). Our first goal is to verify that the anchor function actually dictates the co-location probability of nodes over time. Subsequently, we verify that the set of anchor functions leads to a temporal regularity on a global level.

7.5.1 Impact of the Anchor Function

To show that the anchor function dictates the movement of nodes, we have calculated the probability that a node is located at an anchor for specific instances of time within each simulation period. For mobility models which do not model temporal regularities (e.g. the random waypoint mobility model), this probability is constant for $t_{max} \to \infty$.

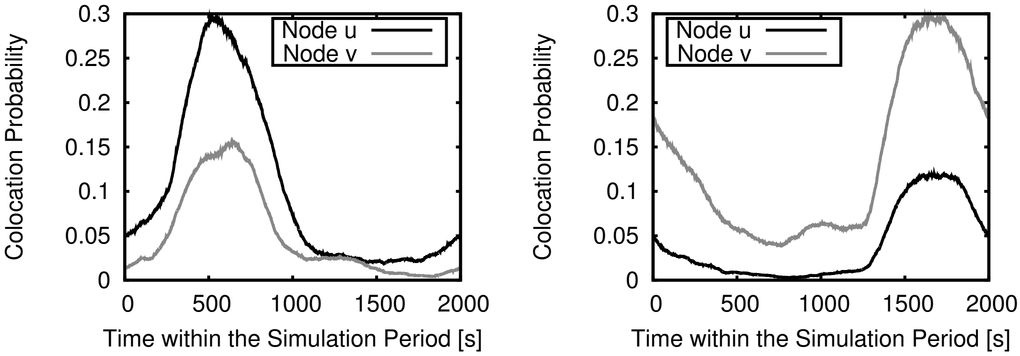
More precisely, we have calculated the co-location probability for an anchor at an arbitrary time $t_c \in \{0, 1, 2, ..t_p - 1\}$ *within* the simulation period, for example at 12am if the simulation period corresponds to 24 hours. Note that in Section 7.4, we have calculated the co-location probability *averaged over the whole simulation duration*, which we used to measure spatial regularities. An approximation of the co-location probability *over time* is the ratio between the number of times a node was located at an anchor at $t \in \{t_c, t_c + t_p, t_c + 2t_p, \ldots\}$ and the total number of simulation periods t_{max}/t_p.

We have analyzed the co-location probability over time for two anchors $a, b \in A$ that are the only home anchors of the nodes u and v respectively, i.e $\mathcal{A}(u) = \{a\}$ and $\mathcal{A}(v) = \{b\}$.



(a) ***Anchor a.*** (b) ***Anchor b.***

Figure 7.14: Anchor function for the home anchors of nodes u and v.



(a) ***Anchor a.*** (b) ***Anchor b.***

Figure 7.15: Co-location probability of two social acquaintances u, v against time.

Furthermore, we defined a social relation between u and v with $w(u, v) = 0.8$. We have observed the co-location probability for several simulations based on this setup using an extended simulation duration of $t_{max} = 10^7 s$ and a short simulation period $t_p = 2000s$. Thus, we executed 5000 simulation periods. We chose these parameters to obtain a representative approximation of the co-location probability. Social relations were set to be active the whole simulation duration. Above described relations between a, b, u and v were fixed for all conducted simulations.

Figure 7.14 shows the anchor function for a and b. Figure 7.15 shows the co-location probability of u and v (y-axis) at both a and b against $t_c \in [0, 1999]$ (x-axis). Note that this is the result of a single simulation run only. However, the qualitative results were the same for all performed simulations. Due to random factors like different social acquaintances, anchor positions, and anchor functions of anchors besides a and b, the exact quantitative characteristics were different for each simulation run.

However, in each measurement we observed two distinct probability peaks that match with the corresponding anchor function. More precisely, the probability that a node is located at its home anchor during the probability peak of the corresponding anchor function is significantly higher (up to 30%) compared to time intervals in which the anchor function is low (below 5% co-location probability). Nodes are also strongly influenced by home anchors of social acquaintances. Furthermore, a certain displacement of the probability peaks is visible. The reason for this is that nodes require some time to react to a sudden increase of the anchor function at time t_0. Most nodes may stay at other anchors at t_0 and will not select a new anchor until their current dwell time is over. Additionally, nodes require some time to travel to an anchor. Respectively, if the anchor function suddenly diminishes, nodes may still stay with the corresponding anchor until they leave by choosing a new anchor. This displacement should be kept in mind while designing an anchor function for a specific location.

As our results suggest, the movement of nodes is dictated by periodic anchor functions. However, it remains to show that these temporal dependencies lead to temporal regularities on a global level. In Section 3.4, we have already introduced the network similarity index (NSI) as a metric to show temporal regularities in an empirical way. Thus, we have calculated the NSI for our simulated mobility traces. Again, we have chosen a simulation period of $t_p = 2000s$ to yield representative results and to enable a feasible calculation of the NSI. All other parameters are generated based on our standard simulation setup.

7.5.2 Global Temporal Regularities

Figure 7.16 shows the NSI (y-axis) for time gaps $t_g \in [700, 7000]$ (x-axis) using different models to generate social networks. We observe a similar qualitative behavior for all social network models. The curve exhibits distinct peaks around $t \in \{2000, 4000, 6000\}$. This corresponds to the chosen simulation period of $t_p = 2000$. Thus, nodes reappear after a period of approximately t_p with increased probability. It is observable that the characteristics shown in Figure 7.16 are very similar compared to the characteristics of real mobility traces, as shown in Figure 3.2 on page 28. Note that the *quantitative* values of the NSI curve are strongly dependent on the number of measured locations.

Since α controls the influence of the location attraction and hence the influence of the individual anchor functions, we evaluated the NSI for different $\alpha \in [0, 0.999]$. The results are shown in Figure 7.17. We observe that an approximately linear dependency between α and the strength of the probability-peaks exists. As α controls the influence of the location attraction, it controls the strength of the temporal regularities. Thus, α represents an important parameter for the evaluation of applications which exploit the habitual behavior of humans. As real traces show, the degree of repetitive behavior is scenario-dependent (cf. Figure 3.2 on page 28). Hence, α also allows the adjustment of the simulation to match different real scenarios.

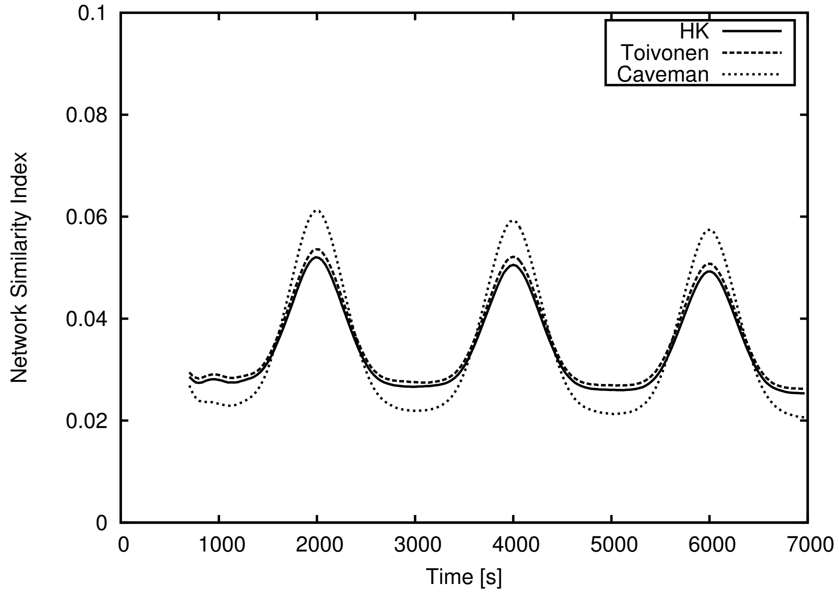


Figure 7.16: NSI for different input social networks.

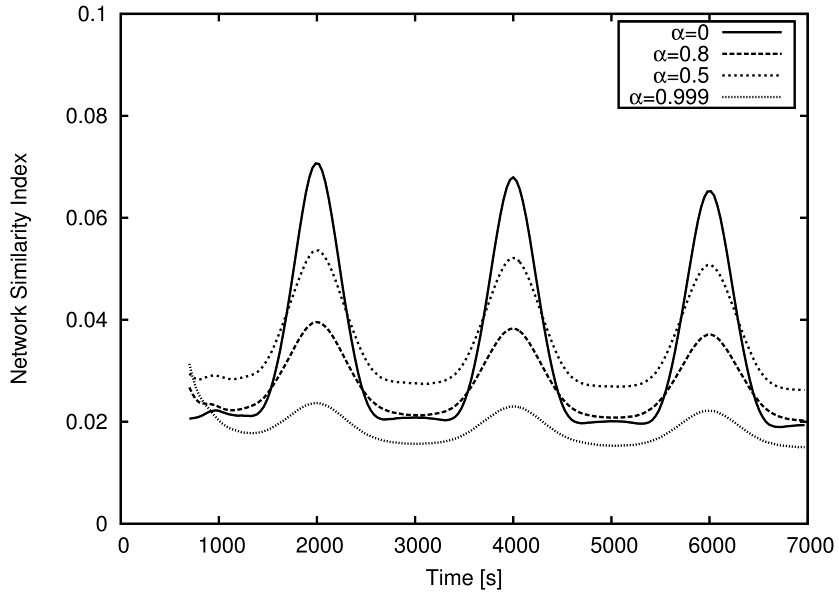


Figure 7.17: NSI against α.

7.6 Group Movement

In this section, we present the evaluation of the probabilistic group mobility created by our model. As discussed in Section 5.3, the probability that a node joins a group is controlled by g_{thresh}. Figure 7.18 shows the impact of g_{thresh} (x-axis) on the number of group movements (i.e. at least two nodes move together) per 1000 simulation ticks (y-axis). We observe a linear decrease of the number of group movements if g_{thresh} is increased until $g_{thresh} = 0.4$. After that, the decrease of the number of group movements proceeds slightly faster. We suspect the following reason for this behavior: After a group of nodes moved together to a new destination anchor and the dwell time of one of the old group members expires, the group or a part of the group may yet again perform a group movement because their dwell time expires at about the same time. This means that if a single group movement is created, it may produce multiple consecutive group movements in the following. However, this self-strengthening effect decreases for $g_{thresh} \rightarrow 1$ because the probability that the old group performs another group movement decreases. This leads to a faster decrease of the number of group movements beyond $g_{thresh} = 0.4$.

We also notice that the number of group movements is significantly increased if the caveman model is used to generate social networks. Once again, we suspect that this may be explained by the increased clustering coefficient. The latter yields an increased probability that three or more nodes are social acquaintances of each other. In other words, a strong clustering increases the number of cliques of size three or greater in the social network. Thus, it is more probable that groups of nodes $V' \subseteq V$ with $|V'| > 2$ meet each other at the same time. This again leads to an increased probability that the dwell time of multiple nodes expires approximately at the same time which translates into the probability to create group movements.

We are also interested if frequent group movements reduce the reflection of the social network. Figure 7.19 shows the impact of g_{thresh} (x-axis) on the reflection metric for critical pairs. Only a small decrease of the reflection is observable if the the number of group movements is increased (by decreasing g_{thresh}).

7.7 Discussion

In this section, we have presented the evaluation of our social mobility model. We have shown that the model reflects the input social network quite well. An optimal reflection of the social network seems to require that the movement of nodes depends on the attraction towards nodes *and* locations. However, a small influence of either type of attraction is enough to yield good results. Furthermore, we have shown that the concept of node repulsion is important to yield a good reflection and realistic inter-contact characteristics.

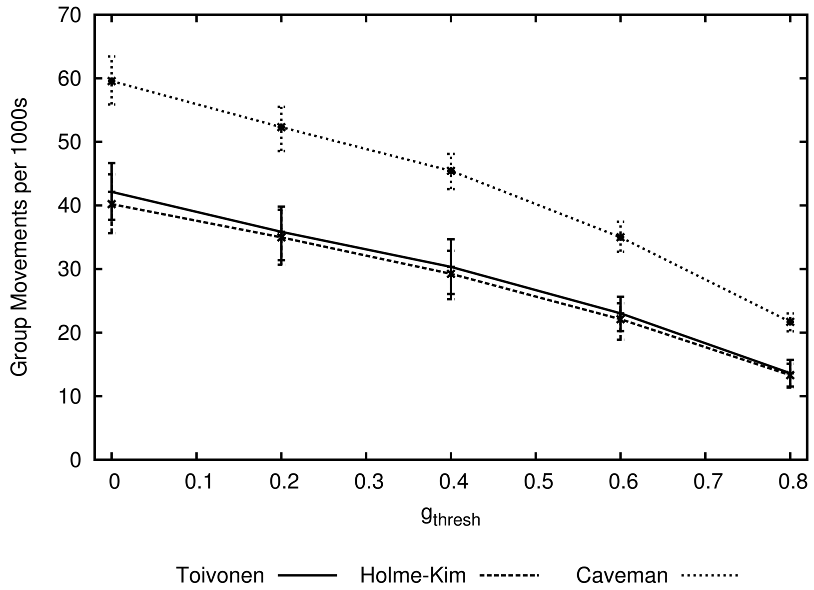


Figure 7.18: Number of group movements against g_{thresh}.

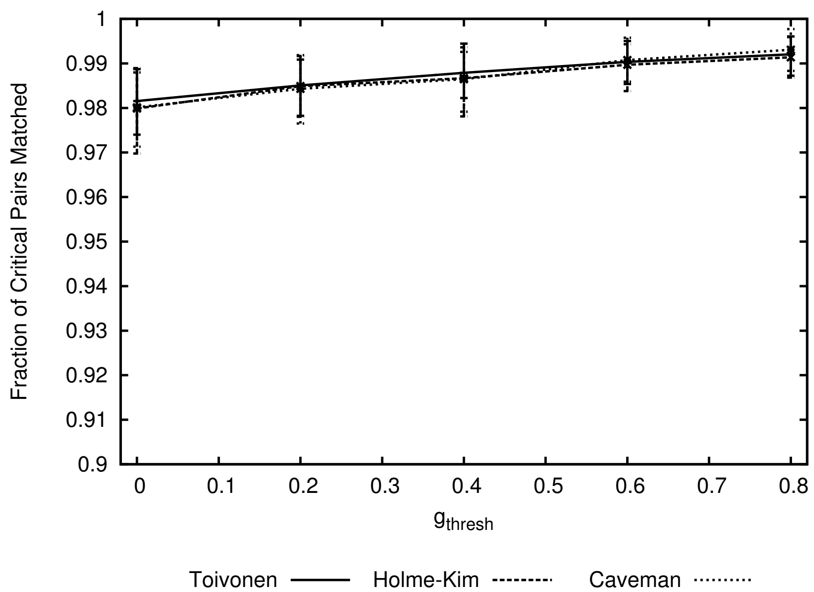


Figure 7.19: Social network reflection against g_{thresh}.

The evaluation has shown that our model captures the dichotomy of real mobility traces quite well. The characteristics exhibited by the temporal and spatial regularties are also very similar to characteristics of empirical results. This confirms the validity of our model.

The fact that the input network is very closely reflected implicates that the structural properties of the social network are reflected as well. For example, the number of meetings with social acquaintances is proportional to its sociability. Thus, a hub in the input network produces many more meetings compared to nodes characterized by a low node degree. Hence, a scale-free node degree in the input network leads to a scale-free distribution of both the number of total meetings between nodes and the average connectivity.

Chapter 8

Advanced Concepts

During the process of creating this thesis, we have developed some ideas to improve the realism of our model even further. Due to time restrictions, we have not implemented these ideas. Nevertheless, we discuss them in the following.

8.1 A Generalized Interpretation of Social Relations

In social networks, a relation may have many different interpretations. For example, an edge between two individuals may define 'has worked together' or 'meet more often than one time per month'. It may refer to a relation between friends, family, or colleague. Besides the nature of the social relation, a number of different factors, like frequency/duration of contact, strong history, connectedness, social capital and many more [Hit03], have to be mapped to a single numerical weight. The concepts shown in Chapter 5 are based on the interpretation that a weight is proportional to the frequency of regular meetings. Such meetings differ from short, random encounters, which typically are not a sign of social relations.

However, other interpretations on the level of mobility are possible. Assume, for example, an individual has frequent encounters with a colleague five days a week. In addition, this individual has a relative whom he visits only once per month, but then for the hole weekend. Based on our described definition, the social relation with the relative is considered significantly less important compared to the social relation with the colleague. However, if such an individual is asked to weight both relationships, he may probably not reflect this interpretation in his answer.

Thus, a social relation of a certain weight may be characterized by frequent and short encounters, or by rather seldom but long encounters. Both types of social relations may be important for mobile applications. Though seldom, long encounters may lead to a long delay, they enable the mobile application to transmit a larger volume of data per contact. Thus, it may be worthwhile to incorporate both types of relations in a social mobility model.

Essentially, if we change the interpretation of a social relation, this means changing the reflection requirement. Instead of creating a number of meetings at anchors proportional to the weight of a social relation, meetings could be produced based on some formula using both the number of meetings and the average meeting duration as a metric. A simple example of such a metric would be to use the *product* between the average meeting duration and number of meetings. Thus 10 meetings of 2 minutes length would be considered equivalent to 5 meetings of 4 minutes length in terms of the weight of a relation. Let $dur(u, v, t)$ denote the average meeting duration of two nodes u, v until time $t \in T$. We propose the following generalization:

Definition 8 (Generalized Reflection Requirement) *For two arbitrary nodes $u, v \in V$, the weighted product*

$$p(v, u, t) := m(v, u, t)^{\epsilon} \cdot dur(v, u, t)^{1-\epsilon}, \ \epsilon \in [0, 1]$$

should be proportional to the strength of their social relation $w(u, v)$ for $t \to \infty$.

The parameter ϵ weights between the frequency and the average duration of meetings between two nodes. Note that the special case of $\epsilon = 1$ yields our original definition of the reflection requirement.

Since the correction factor (see Section 5.5) adjusts the social attraction between two nodes to satisfy the reflection requirement, we have to adapt the calculation of the meeting-quota to our generalized definition. Until now, the total number of meetings $M(v, t)$ of a node v until simulation tick t with active social acquaintances was essentially redistributed among the active social relations according to a perfect reflection, to yield the meeting-quotas. To satisfy our new interpretation, we have to replace the number of meetings between two nodes $m(u, v, t)$ with our weighted product $p(u, v, t)$. Thus, Equation (5.4) on page 62 must be replaced by

$$M(v, t) := \sum_{u \in C_{active}(v,t)} p(v, u, t).$$

This adjusts the definition of the social influence. Now, the accumulated weighted product p of the node with the smallest accumulated weighted product yields our new reference social influence, which is 'redistributed' among the active social relations to yield the meeting-quota. Because all other formulas, including the calculation of the isolation probability i.e. Equation (5.7), are based on the definition of $M(v, t)$, no further changes are necessary. Thus, by making this simple change, our model adjusts the social attraction between two nodes according to the new generalized definition of the reflection requirement.

8.2 Towards a Trace-based Social Mobility Model

Many parameters of our model are dependent on random generation. Furthermore, in many cases there exists no advanced model that may be employed to create such data based on real characteristics. For example, there are researched models for generating social networks. However, we have used a uniform generation of the weights of social relations due to the lack of an appropriate model. To provide for further realism, we introduce the idea of using data extracted from real mobility traces to fill the gap. This moves our model towards a trace-based approach (cf. Section 2.4.2). We have identified the following parameters that may be extracted from mobility traces:

- **Input Network.** A mapping-function dependent on parameters like the number of contacts, total contact duration, or inter-contact duration may be used to extract a weight for each pair of nodes. This function should depend on the interpretation of social relations. For example, if the original reflection requirement is used, the *number of contacts* between two nodes should be mapped to a weight. Because seldom contacts between nodes without a social relation are quite common, a threshold for the number of contacts should be employed to infer an actual social relation. Alternatively, since our model produces *long* meetings at anchors, it may be sensible to count only the number of contacts with a length above a threshold.

  There are other sources to retrieve an input network. Sociologists use questionnaires or interviews to create an actual social network [Sco00]. In addition, existing databases could be utilized. For example, online social networks like Facebook [Fac] may be used to extract potentially very large social networks (or subsets). Weights of the social relations may be assigned by statistical data like the number of messages sent to each other, bulletin board entries, or profile data. However, there may be privacy issues involved.

- **Active Relation Intervals.** Periodic time intervals in which nodes have a high probability to be in contact could be extracted from real mobility traces. We sketch the following possible approach:

  First, the length of the period must be defined, e.g. a day or a week. Then, we calculate the contact probability for each pair of nodes within this period, which is very similar to the calculation of the co-location probability over time (see Section 7.5). For each pair of nodes, we count the number of times in which the nodes have been in contact at a discrete set of points within the period (e.g. at 13:23pm using a period of 24 hours). A possible approach would be to use a minute as a discrete time step. The number of contacts at each periodic instance of time t is then normalized by the number of measured periods to yield the approximated contact probability at t. Finally, all time intervals within the period are identified in which the number of measured contacts is above a threshold p_{thresh}. We then use χ with $t \in \chi(e) :\Leftrightarrow p_{contact}(e,t) > p_{thresh}$ for each pair of nodes e as input for our model.

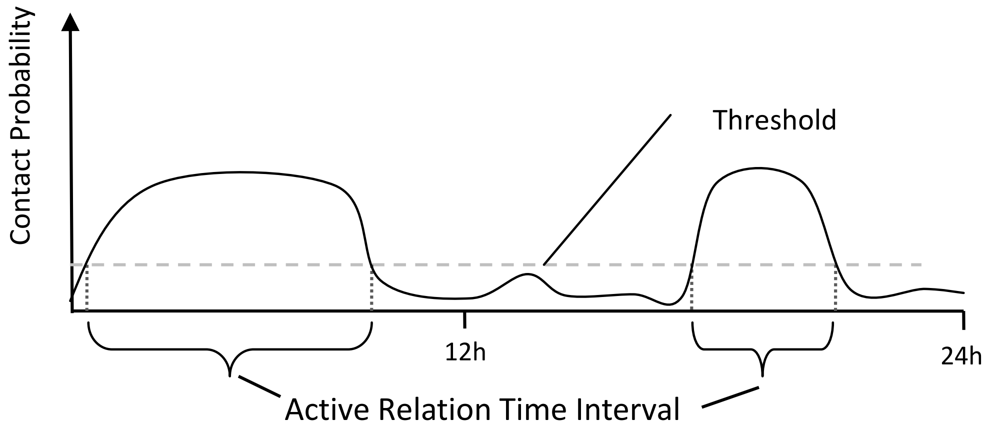


Figure 8.1: Example for the extraction of time intervals in which a social relation is active.

Figure 8.1 shows a constructed example for the measured contact probability (y-axis) of two nodes over the time within the period of one day (x-axis). Two distinct probability peaks are visible. The dashed line represents the chosen threshold. Time intervals, in which the contact probability is above the threshold, represent the time intervals in which the social relation is active.

- **Geo-social Patterns.** If the trace is enriched with location information, like GPS-traces, it is possible to identify the social sphere of individuals. Some existing work already focuses on this [GKBB09, BGK+07]. For example, Adams et al. propose a method to extract significant locations for each user, collected by a mobile, wearable device [APV08, APV06]. They apply a clustering algorithm to a set of GPS coordinates which yields for each node the locations of the social sphere. They also identify locations which are shared among nodes and infer social relations. This information may then be used to create home anchors of different nodes. For example, a location could be associated as a home anchor to the node with the most frequent visits. In addition, corresponding social relations could be created based on the strength and/or number of shared locations between two nodes. The authors also identify the convex hulls of the extracted locations, which could be used to specify the geographical boundaries of the anchors in our model.
- **Anchor Function.** If GPS or AP-based traces are used, it is possible to approximate the co-location probability (cf. Section 7.5.1) at specific locations over a set of discrete points in time within a period (e.g. a day). Based on this probability distribution, a corresponding anchor function could be created.
- **Dwell Time Distributions.** In Section 2.4.2, we have introduced mobility models that extract the pause time distribution for each AP using an AP-based trace. The same

principle may be applied to yield dwell time distributions for each anchor corresponding to an AP. Anchors may even be placed according to coordinates associated with an AP. Instead of using AP-based traces, one could also employ mobility traces which feature GPS coordinates. For example, the durations an individual remains within the boundary of a identified convex hull of a significant location may be used to create a dwell time distribution for the anchor which represents the location.

- **Speed Distributions.** GPS trace data allows one to extract a speed distribution for different users, similar to existing mobility models [KKK06].

Chapter 9

Conclusions

9.1 Summary

The goal of this thesis was to create a realistic mobility model, which focuses on the simulation of social context. Our proposed model uses a weighted social network as input to generate its mobility scheme. The basic assumption is that individuals, which share a social relation, tend to meet more frequently and regularly, in contrast to individuals without a social relation. Based on this assumption, the model reflects the social network, i.e. produces meetings between nodes with a probability which is proportional to the weight of their social relation.

To create a realistic mobility model, we have identified several characteristics of human mobility that are of importance for the evaluation of mobile applications. Besides explicitly modeling (active) social relations, we have advocated that a realistic mobility model should capture the characteristics of temporal and spatial regularities as well as group mobility. Based on these requirements, we have proposed a social mobility model.

Basically, mobile nodes move between a set of anchors, which represent abstract locations. The movement of nodes is dictated by the attraction towards certain abstract locations and the attraction towards social acquaintances. Some nodes may be repulsed by others in order to yield a good reflection of the input network. The anchor attraction is based on a periodic characteristic anchor function, which creates a temporal regularity. In addition, mobile nodes are only attracted to a subset of all anchors, which we call the social sphere of a node. This creates a spatial regularity. Our model also generates probabilistic group movements based on a parameterized probability.

We have proposed a concept to yield a good reflection of the social network, despite 'disturbances' introduced by different structural properties of social networks, concepts of the geographical mobility model, and a heterogeneous node degree.

We have discussed several integration points for geographical mobility models and shown why this integration does not prevent a proper reflection of the social network. Furthermore, it was pointed out that our model generalizes several existing (social) mobility models.

The results of our evaluation have shown that the model reflects the input social network with an accuracy of over 99% for the standard parameter setup. We have also shown that our model exhibits characteristics of spatial and temporal regularities that are very similar compared to existing empirical results. In particular, our model allows parameterizing the degree of repetitive behavior, which is important to evaluate applications which feature mobility prediction. Additionally, the inter-contact distribution of our model captures the dichotomy measured in real mobility traces quite well. We have used several different models to generate social networks and thus indicated that these results are independent of the used input network.

We have discussed how a more generalized interpretation of social relations may be integrated into our mobility model. In addition, we have sketched how our model may be enriched by data extracted from real mobility traces, which may increase the degree of realism.

9.2 Contributions

In this thesis, we have developed a social mobility model which reflects an *arbitrary* input network. This may enable the research and the evaluation of mobile applications which exploit either the structure of social networks or knowledge about social context. In addition, it allows the evaluation of mobile applications which focus on exploiting group movement or mobility prediction mechanisms. This is the first mobility model that combines the reflection of an arbitrary social network with spatial and temporal regularities. The fact that our model captures characteristics observed in real traces confirms the validity of the model.

We consider the primary contribution of our work to be the generalizational character of the proposed model. We have shown (cf. Section 5.7) that our model can be specialized to yield several existing social mobility models. Furthermore, the model allows the integration of geographical concepts such as obstacles and speed distributions, and thus may enrich the sociological aspects with further realism. The fact that our model reflects an arbitrary social network has many advantages that we will discuss in the following.

Although some properties of social networks seem to be omnipresent, other properties are more or less pronounced in different social networks. Our social mobility model allows the evaluation based on social networks exhibiting different structural properties. This enables to research the influence of different structural properties on the performance of mobile applications without having to change the simulation model. If the exact properties of the target social network are known (e.g. because a trace exists) the mobility scheme of our model may be tailored to this specific scenario, either by directly using a specific social network or by generating networks with the desired structural properties. It is also possible to integrate networks extracted from a real mobility trace (cf. Section 8.2). However, even if the researcher does not know the exact structural properties of his target social network, the application may be evaluated using multiple social networks with realistic structural properties (cf. Section 2.2). If the application performs well in each case, this can be considered as a further indicator for the validity of the evaluated concept.

One may argue that it would be possible for a researcher to evaluate his system using different mobility models, for example, one that does reflect the scale-free structure of many social networks and one that rather reflects a cut-off. However, different models create mobility in different ways and therefore exhibit other properties such as inter-contact time characteristics. Therefore, if the researcher measures a different result in different mobility models, he cannot say if this is due to the different properties of the mobility model or due to the different structure of the underlying social network.

The research on complex social networks is still an ongoing process. We expect that in the future new structural properties of social networks are discovered. Our model allows to incorporate such results in its mobility scheme without changing the model or even the implementation. This enables the fast exploration of the use of such properties without the need to develop an appropriate mobility model first.

The research of the relation between the mobility patterns of individuals (which defines the connectivity graph) and the underlying social network is still in a very early stage and only a few results are known. Based on these few results (as discussed in Section 3.1), we have proposed a simple interpretation for a social relation defined by the reflection requirement. We have also discussed how to generalize this interpretation (see Section 8.1). However, we do not know which interpretation may be considered to be realistic. To actually choose a concrete interpretation, one has to find out how exactly a social network should be mapped to a measure of the frequency/length of encounters between the individual users. Note that our interpretation further assumes that if two individuals do not share a social relation, they only meet by random encounters.

However, as future research progresses, results in this area may be easily integrated into our social mobility model. First, we could choose an actual interpretation for social relations. Second, the generated social network can be adapted to reflect these results. The basic assumption of the proposed model is that a social network N_s is directly mapped to the mobiliy scheme and thus to a connectivity graph N_c. However, if a less obvious mapping between N_s and N_c is known in the future, N_s could be transformed to a network N_s' that incorporates this mapping. This transformed network may then be provided as the actual input for our model. For example, consider the hypothetical case that investigations discover that two-hop friends (i.e. nodes with a distance of 2 in the social network), typically meet with a specific probability, despite having no social relation. Intuitively, we believe that this might be true. In this case, the social network N_s can simply be mapped to N_s' in which every two-hop friend has a social relation of a certain weight.

Of course, one may avoid the problem of interpreting the mapping between a social network and the mobility scheme altogether, by using an input network, extracted from real traces. In this case, N_s corresponds to N_c. However, because such traces are only sparsely available, this method may *currently* yield only a few input networks with a small number of nodes. Thus, for an empirical evaluation, the random generation according to a *realistic* model may be more desirable.

9.3 Limitions

The generalizational character of our model comes at a price. Compared to many other mobility models, our approach is rather complex. This may lead to difficulties in determining its properties analytically. However, we have shown that this complexity does not lead to an inefficiency, i.e. our model still generates large mobility traces with a significant number of nodes in acceptable time. It remains for further work to show if a simpler model can be found that still satisfies all requirements we have set for our model.

As we have shown in the evaluation, the number of group movements per time is controlled by a single parameter. However, the more group movements should take place, the greater the probability that a node does not stay for its whole dwell time at an anchor (cf. Section 5.3). Thus, if a scenario characterized by many group movements should be captured, nodes do not obey the behavior specified by the dwell time distribution. However, we consider the dwell time distribution to be a significant parameter to adapt the model to specific scenarios. It remains for future work to evaluate if it is possible to integrate both frequent group movements and a dwell time of nodes according to a predefined distribution. One possible approach would be to transform the *desired* dwell time distribution $\tilde{D}_a$ to the actual input dwell time distribution D_a such that the dwell time distribution of anchor a, created due to produced group movements, matches $\tilde{D}_a$. Of course, the mean value of $\tilde{D}_a$ is less than the mean value of D_a. The decision if a node v located at a should join a group movement, could be made by comparing the dwell times produced by a in the past with $\tilde{D}_a$. Of course, the decision should also depend on the desired number of group movements. Thus, the actual created dwell time distribution is shaped to yield $\tilde{D}_a$. This approach may also allow to create a fixed (controlled by a parameter) number of group movements per time, independent of the used input network (cf. Section 7.6). We leave it to future work to show if such an approach is possible.

9.4 Outlook

In our social mobility model, time is structured into multiple simulation periods of fixed length t_p. This yields habitual behavior with a periodicity of t_p. However, in the real world, multiple *hierarchical* simulation periods of different periodicity exist. For instance, an office worker may reappear with high probability at the bus stop for work with a *period of one day*. This individual has the position of a team leader and meets with his team one time per week in a meeting room. Thus, he reappears at this meeting room with high probability after a *period of one week*. In addition, he has a team leader meeting once per month. Therefore, he meets with other team leaders with a *period of one month*. Predictive mobile application may keep track of these hierarchical temporal regularities and may use them to predict connectivity. Thus, it may be sensible to integrate this behavior in a social mobility model. Such hierarchical regularities may be incorporated in our model by providing an anchor function that defines probability peaks with multiple periodicities. A possible approach could be to create an anchor function for each periodicity and to calculate the 'superposition' of the individual anchor functions.

In reality, an individual moves only between a subset of its social sphere at any given time. For example, our office worker would spend the time on weekdays during the day between a set of office rooms. At weekends and in the afternoon, he may spend his time at home and other places, related to his recreational time. The existence of such so called mobility profiles, i.e. a set of locations which are important within a certain (periodic) time interval, has been confirmed in real mobility traces [GBNQ06]. Mobile applications may collect and exploit knowledge about the currently active mobility profile. Note that our model already implicitly creates very simple mobility profiles, created by the concept of active social relations: A node moves only to home anchors of other nodes, to which a social relation exists. However, we suspect that this simplification of mobility profiles is not realistic enough. Future work could investigate the explicit specification and integration of different mobility profiles into our social mobility model.

We have shown several integration points for geographical mobility models (cf. Section 5.6). The next step would be to actual perform such an integration. We expect that a compound model improves the realism of the model even further. It might also be interesting to evaluate the reflection of the social network in the compound model. Note that we have already tried to simulate the corresponding effects during the evaluation of our model (cf. Section 7.2.4).

Currently, only very few direct contact mobility traces are publicly available which feature many nodes and at the same time a representative number of contacts. If more such traces become available in the future, it might be interesting to compare their characteristics to characteristics of our social mobility model, like the influence of the clustering coefficient on the mobility behavior. In general, if further omnipresent characteristics of human mobility, such as the dichotomy of the inter-contact distribution are discovered, a comparison may be insightful as well.

Many parameters of our proposed model are randomly generated. Alternatively, we have shown how to extract these parameters from real mobility traces. However, if a researcher wants to customize the simulation to a concrete target scenario, many parameters have to be adjusted. The custom setting of the parameters may be a time consuming event. However, for a detailed and realistic evaluation of the target scenario it may be worth the effort. Thus, to assist in the customization of the parameters of our model, we envision the development of a graphical tool. Features of such a tool may include the following:

- Visualization of the simulation area and placements of the anchors. The position of individual anchors could be changed by drag&drop functionality. A map could be imported to place as an underlay behind the simulation area. This helps the scenario designer to place the anchors at corresponding locations.
- Setting of the anchor function according to a list of predefined models.
- Selecting the dwell time distribution from different parameterized distributions such as Gaussian, uniform, Poisson, etc.
- Assignment of home anchors to nodes and visualization of the individual social sphere of a selected note.

The detailed design of such a graphical tool remains open for future work.

Bibliography

[AB02] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002. (Cited on pages 9 and 12)

[ACL00] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180, 2000. (Cited on page 9)

[ALPH01] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman. Search in power-law networks. *The American Physical Society*, 64:46135–46143, 2001. (Cited on page 9)

[APV06] Brett Adams, Dinh Phung, and Svetha Venkatesh. Extraction of social context and application to personal multimedia exploration. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 987–996, 2006. (Cited on page 106)

[APV08] Brett Adams, Dinh Phung, and Svetha Venkatesh. Sensing and using social context. *ACM Trans. Multimedia Comput. Commun. Appl.*, 5(2):1–27, 2008. (Cited on pages 6, 16, 26 and 106)

[ASBS00] LAN Amaral, A. Scala, M. Barthelemy, and HE Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, pages 11149–11152, 2000. (Cited on page 10)

[ATB08] A. Aravind, H. Tahir, and Baldeep. Realmobisim: realistic mobility simulator and analyzer. In *PM2HW2N '08: Proceedings of the 3nd ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks*, pages 186–189, 2008. (Cited on page 20)

[BA99] A.L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. (Cited on pages 9 and 11)

[BCP07] Chiara Boldrini, Marco Conti, and Andrea Passarella. Users mobility models for opportunistic networks: the role of physical locations. In *IEEE Wireless Rural and Emergency Communications - WRECOM07*, 2007. (Cited on page 33)

[BCP08] Chiara Boldrini, Marco Conti, and Andrea Passarella. Contentplace: social-aware data dissemination in opportunistic networks. In *MSWiM '08: Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, 2008. (Cited on page 15)

[BGK+07] Cristian Borcea, Ankur Gupta, Achir Kalra, Quentin Jones, and Liviu Iftode. The mobisoc middleware for mobile social computing: challenges, design, and early experiences. In *MOBILWARE '08: Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications*, pages 1–8, 2007. (Cited on page 106)

[BJN+02] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *PHYSICA A*, 311:590–614, 2002. (Cited on page 9)

[BLdAF06] Vincent Borrel, Franck Legendre, Marcelo Dias de Amorim, and Serge Fdida. Simps: Using sociology for personal mobility. *CoRR*, abs/cs/0612045, 2006. (Cited on page 36)

[BLM+06] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006. (Cited on pages vii, 7 and 14)

[BRS+04] D. Bhattacharjee, A. Rao, C. Shah, M. Shah, and A. Helmy. Empirical modeling of campus-wide pedestrian mobility: Observations on the usc campus. In *Proceedings of the IEEE Vehicular Technology Conference*, volume 4, pages 2887– 2891, 2004. (Cited on page 21)

[BSH03] F. Bai, N. Sadagopan, and A. Helmy. The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks. *Ad Hoc Networks*, 1(4):383–403, 2003. (Cited on pages 20 and 23)

[Cai07] H. Cai. Crossing over the bounded domain: from exponential to power-law intermeeting time in manet. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 159–170, 2007. (Cited on page 18)

[CBD02] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483–502, 2002. (Cited on pages 17, 20, 21 and 23)

[CCJ90] Brent N. Clark, Charles J. Colbourn, and David S. Johnson. Unit disk graphs. *Discrete Math.*, 86(1-3):165–177, 1990. (Cited on page 79)

[CFL08] Augustin Chaintreau, Pierre Fraigniaud, and Emmanuelle Lebhar. Opportunistic spatial gossip over mobile social networks. In *WOSP '08: Proceedings of the first workshop on Online social networks*, pages 73–78, 2008. (Cited on page 67)

[CHC+05] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. Technical report, UCAM-CL-TR-617, University of Cambridge, 2005. (Cited on pages vi and 18)

[CHD+07] Augustin Chaintreau, Pan Hui, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007. (Cited on pages 17, 18 and 93)

[CHK05] G. Chen, H. Huang, and M. Kim. Mining frequent and periodic association patterns. Dartmouth College, Coomputer Science and Engineering. Technical report, TR 2005-550, University of Cambridge, 2005. (Cited on page 27)

[CJGS08] Rubén Cuevas, Eva Jaho, Carmen Guerrero, and Ioannis Stavrakakis. Onmove: a protocol for content distribution in wireless delay tolerant networks based on social information. In *CONEXT '08: Proceedings of the 2008 ACM CoNEXT Conference*, pages 1–2, 2008. (Cited on page 15)

[CLF07] Vania Conan, Jérémie Leguay, and Timur Friedman. Characterizing pairwise inter-contact patterns in delay tolerant networks. In *Autonomics '07: Proceedings of the 1st international conference on Autonomic computing and communication systems*, pages 1–9, 2007. (Cited on page 25)

[CMMP08] Paolo Costa, Cecilia Mascolo, Mirco Musolesi, and Gian Pietro Picco. Socially-aware Routing for Publish-Subscribe in Delay-tolerant Mobile Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications (JSAC)*, 26(5):748–760, 2008. (Cited on page 15)

[Dar] Dartmouth College. CRAWDAD. `http://crawdad.cs.dartmouth.edu/`. (Cited on page 17)

[Dav00] V. Davies. Evaluating mobility models within an ad hoc network. Master's thesis, Colorado School of Mines, 2000. (Cited on page 20)

[DCSJS08] Munmun De Choudhury, Hari Sundaram, Ajita John, and Dorée Seligmann. Dynamic prediction of communication flow using social context. In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 49–54, 2008. (Cited on page 6)

[DDGDA05] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 9:P09008, 2005. (Cited on page 10)

[DH07] Elizabeth M. Daly and Mads Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pages 32–40, 2007. (Cited on page 15)

[DMW03] P.S. Dodds, R. Muhamad, and D.J. Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003. (Cited on page 8)

[DNS08] Djamel Djenouri, Elmalik Nekka, and Wassim Soualhi. Simulation of mobility models in vehicular ad hoc networks. In *SOMITAS '08: Proceedings of the 2008 Ambi-Sys workshop on Software Organisation and MonIToring of Ambient Systems*, pages 1–7, 2008. (Cited on page 20)

[DYB03] G.F. Davis, M. Yoo, and W.E. Baker. The small world of the American corporate elite, 1982-2001. *Strategic Organization*, 1(3):301–326, 2003. (Cited on page 10)

[Ein56] A. Einstein. *Investigations on the Theory of the Brownian Movement.* Dover Publications, 1956. (Cited on page 20)

[EKKO08] Frans Ekman, Ari Keränen, Jouni Karvo, and Jörg Ott. Working day movement model. In *MobilityModels '08: Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 33–40, 2008. (Cited on page 3)

[EP05] N. Eagle and A. Pentland. Social serendipity: mobilizing social software. *Pervasive Computing, IEEE*, 4(2):28–34, 2005. (Cited on page 16)

[EP06] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006. (Cited on pages 17 and 18)

[ER59] P. Erdos and A. Renyi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959. (Cited on page 8)

[Fac] Facebook Inc. Facebook. `http://www.facebook.com`. (Cited on page 105)

[Fal03] Kevin Fall. A delay-tolerant network architecture for challenged internets. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 27–34, 2003. (Cited on page 14)

[FS64] TJ Fararo and M.H. Sunshine. *A study of a biased friendship net.* Youth Development Center, Syracuse University, 1964. (Cited on page 10)

[GBNQ06] Joy Ghosh, Matthew J. Beal, Hung Q. Ngo, and Chunming Qiao. On profiling mobility and predicting locations of wireless users. In *REALMAN '06: Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality*, pages 55–62, 2006. (Cited on pages 27 and 113)

[GDDG+03] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in organisations. *Physical Review E*, 68:65103–65106, 2003. (Cited on page 10)

[GGP08] Sabrina Gaito, Giuliano Grossi, and Federico Pedersini. A two-level social mobility model for trace generation. In *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 457–458, 2008. (Cited on pages 3, 35 and 37)

[GHB08] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008. (Cited on pages 26, 28 and 66)

[GKBB09] A. Gupta, A. Kalra, D. Boston, and C. Borcea. MobiSoC: a middleware for mobile social computing applications. *Mobile Networks and Applications*, 14(1):35–52, 2009. (Cited on page 106)

[GN02] Michelle Girvan and M. E. J. Newman. Community structure in social and biological networks. *PROC.NATL.ACAD.SCI.USA*, 99:7821–7826, 2002. (Cited on page 10)

[Gra73] M.S. Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973. (Cited on pages 9 and 25)

[HBE+01] J. Heidemann, N. Bulusu, J. Elson, C. Intanagonwiwat, K. Lan, Y. Xu, W. Ye, D. Estrin, and R. Govindan. Effects of detail in wireless network simulation. *Proceedings of the SCS Multiconference on Distributed Simulation*, pages 3–11, 2001. (Cited on page 23)

[HCS+05] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, 2005. (Cited on pages 14 and 17)

[HCY08] Pan Hui, Jon Crowcroft, and Eiko Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 241–250, 2008. (Cited on page 15)

[HD03] RA Hill and RIM Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, 2003. (Cited on page 10)

[Her03] Klaus Herrmann. Modeling the sociological aspects of mobility in ad hoc networks. In *MSWIM '03: Proceedings of the 6th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, pages 128–129, 2003. (Cited on pages 20, 31 and 37)

[HFB08] J. Haerri, F. Filali, and C. Bonnet. Mobility models for vehicular ad hoc networks: a survey and taxonomy. Technical report, RR-06-168, Institut Eurecom Department of Mobile Communications, 2008. (Cited on page 20)

[HGPC99] Xiaoyan Hong, Mario Gerla, Guangyu Pei, and Ching-Chuan Chiang. A group mobility model for ad hoc wireless networks. In *MSWiM '99: Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, pages 53–60, 1999. (Cited on page 21)

[HH05] Wei-jen Hsu and Ahmed Helmy. IMPACT: Investigation of Mobile-User Patterns Across University Campuses using WLAN Trace Analysis. CoRR abs/cs/0508009, 2005. (Cited on pages 28 and 93)

[HH06a] Wei J. Hsu and Ahmed Helmy. On nodal encounter patterns in wireless LAN traces. In *Proceedings of the Second Workshop on Wireless Network Measurements (WiNMee 2006)*, 2006. (Cited on pages 8 and 17)

[HH06b] Wei-jen Hsu and A. Helmy. On modeling user associations in wireless lan traces on university campuses. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, pages 1–9, 2006. (Cited on pages vi, 27, 28 and 36)

[Hit03] J.M. Hite. Patterns of multidimensionality among embedded network ties: a typology of relational embeddedness in emerging entrepreneurial firms. *Strategic Organization*, 1(1):9–49, 2003. (Cited on page 103)

[HK02] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65(2):026107–026112, 2002. (Cited on page 12)

[HMS+05] Wei-jen Hsu, Kashyap Merchant, Haw-wei Shu, Chih-hsin Hsu, and Ahmed Helmy. Weighted waypoint mobility model and its impact on ad hoc networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(1):59–63, 2005. (Cited on page 21)

[HSPH07] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling Time-variant User Mobility in Wireless Mobile Networks. In *Proceedings of IEEE INFOCOM*, 2007. (Cited on page 34)

[HSPH08] Wei-jen Hsu, Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Ahmed Helmy. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *Computing Research Repository*, abs/0810.3935, 2008. (Cited on page 34)

[Hsu08] Wei-jen Hsu. *Behavioral Analysis, User Modeling, and Protocol Design based on Large-scale Wireless Network Traces.* PhD thesis, University of Florida, 2008. (Cited on pages 34 and 37)

[Hui07] Pan Hui. People are the network: experimental design and evaluation of social-based forwarding algorithms. Technical report, UCAM-CL-TR-713, University of Cambridge, 2007. (Cited on page 24)

[IC09] S. Ioannidis and A. Chaintreau. On the strength of weak ties in mobile social networks. In *in Proceedings of the Second ACM Workshop on Social Network Systems (SNS)*, 2009. (Cited on page 25)

[JBRAS03] Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, and Subhash Suri. Towards realistic mobility models for mobile ad hoc networks. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 217–229, 2003. (Cited on pages 20 and 23)

[JGN01] E.M. Jin, M. Girvan, and MEJ Newman. Structure of growing social networks. *Physical Review E*, 64(4):46132–46135, 2001. (Cited on page 10)

[JM96] David B. Johnson and David A. Maltz. Dynamic source routing in ad hoc wireless networks. In *Mobile Computing*, pages 153–181, 1996. (Cited on page 20)

[JS09] E. Jaho and I. Stavrakakis. Joint interest- and locality-aware content dissemination in social networks. In *Sixth Annual Conference on Wireless On demand Network Systems and Services, IFIP/IEEE WONS 2009*, 2009. (Cited on page 27)

[KBV07] Thomas Karagiannis, Jean-Yves Le Boudec, and Milan Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 183–194, 2007. (Cited on page 18)

[KJRN05] Nishan C. Karunatillake, Nicholas R. Jennings, Iyad Rahwan, and Timothy J. Norman. Argument-based negotiation in a social context. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1331–1332, 2005. (Cited on page 6)

[KKK00] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and inference problems for temporal networks. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 504–513, 2000. (Cited on page 19)

[KKK06] M. Kim, D. Kotz, and S. Kim. Extracting a mobility model from real user traces. In *Proceedings of the 25th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Barcelona, Spain, April*, 2006. (Cited on pages 20, 21, 66, 93 and 107)

[KPR06] T. Kathiravelu, A. Pears, and N. Ranasinghe. Connectivity Models: A New Approach to Modeling Contacts in Opportunistic Networks. In *Proceedings of the 8th international information technology conference (IITC2006)*, 2006. (Cited on page 20)

[LDS04] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. *Lecture Notes in Computer Science*, pages 239–254, 2004. (Cited on page 14)

[LEA+01] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001. (Cited on page 9)

[LH99] B. Liang and ZJ Haas. Predictive distance-based mobility management for PCS networks. In *IEEE INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, volume 3, 1999. (Cited on page 20)

[MCL+08] A. Mtibaa, A. Chaintreau, J. LeBrun, E. Oliver, A.K. Pietilainen, and C. Diot. Are you moved by your social network application? In *Proceedings of the first workshop on Online social networks*, pages 67–72, 2008. (Cited on pages 16, 19 and 24)

[MGC+07] A.G. Miklas, K.K. Gollu, K.K.W. Chan, S. Saroiu, K.P. Gummadi, and E. De Lara. Exploiting social interactions in mobile systems. *Lecture Notes in Computer Science*, 4717:409–428, 2007. (Cited on pages 14, 16 and 36)

[MHM04] Mirco Musolesi, Stephen Hailes, and Cecilia Mascolo. An ad hoc mobility model founded on social network theory. In *MSWiM '04: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 20–24, 2004. (Cited on page 32)

[Mil67] Stanley Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967. (Cited on page 8)

[MM06] Mirco Musolesi and Cecilia Mascolo. A Community based Mobility Model for Ad Hoc Network Research. In *Proceedings of the 2nd ACM/SIGMOBILE International Workshop on Multi-hop Ad Hoc Networks: from theory to reality (REALMAN'06)*, 2006. (Cited on pages 32 and 37)

[MM07] Mirco Musolesi and Cecilia Mascolo. Designing mobility models based on social network theory. *ACM SIGMOBILE Mobile Computing and Communication Review*, 11(3):59 – 70, 2007. (Cited on page 32)

[MM08] M. Musolesi and C. Mascolo. Mobility Models for Systems Evaluation. *Middleware for Network Eccentric and Mobile Applications*, pages 43–62, 2008. (Cited on pages 19, 20 and 21)

[MT07] Kristijan Mihalic and Manfred Tscheligi. 'divert: mother-in-law': representing and evaluating social context on mobile devices. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 257–264, 2007. (Cited on page 6)

[New01] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98:404–409, 2001. (Cited on page 8)

[New02] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20):208701–208704, 2002. (Cited on pages vii, 10 and 14)

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003. (Cited on page 7)

[New06] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103(23):8577–8582, 2006. (Cited on page 10)

[NG04] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113–026127, 2004. (Cited on page 32)

[NP03] MEJ Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):36122–36129, 2003. (Cited on page 10)

[NS2] The network simulator - ns-2. `http://www.isi.edu/nsnam/ns/`. (Cited on page 73)

[NWS02] M. E. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proc Natl Acad Sci U S A*, 99 Suppl 1:2566–2572, 2002. (Cited on page 10)

[NYBK06] T. Nicolai, E. Yoneki, N. Behrens, and H. Kenn. Exploring social context with the wireless rope. *LECTURE NOTES IN COMPUTER SCIENCE*, 4277:874–883, 2006. (Cited on page 29)

[PDFV05] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005. (Cited on page 10)

[PHO02] DD Perkins, HD Hughes, and CB Owen. Factors affecting the performance of ad hoc networks. In *IEEE International Conference on Communications, 2002. ICC 2002*, volume 4, 2002. (Cited on page 23)

[PK78] I. Pool and M. Kochen. Contacts and influence. *Social Networks*, 1:5–51, 1978. (Cited on page 8)

[POL+08] A.K. Pietilainen, E. Oliver, J. LeBrun, G. Varghese, J. Crowcroft, and C. Diot. Experiments in mobile social networking. Technical report, CR-PRL-2008-02-0003, Thomson, 2008. (Cited on page 16)

[PPC06] L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE Communications Magazine*, 44(11):134–141, 2006. (Cited on page 14)

[Rap57] A. Rapoport. Contribution to the theory of random and biased nets. *Bulletin of mathematical biology*, 19(4):257–277, 1957. (Cited on page 9)

[RCC+04] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, 101(9):2658–2663, 2004. (Cited on page 10)

[Rey87] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34, 1987. (Cited on page 21)

[SA03] R. Schlich and K.W. Axhausen. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30(1):13–36, 2003. (Cited on page 28)

[SBGL08] Marc Smith, Vladimir Barash, Lise Getoor, and Hady W. Lauw. Leveraging social context for searching social media. In *SSM '08: Proceeding of the 2008 ACM workshop on Search in social media*, pages 91–94, 2008. (Cited on page 6)

[Sco00] J. Scott. *Social network analysis: A handbook.* Sage, 2000. (Cited on pages 5 and 105)

[SHR05] Illya Stepanov, Daniel Herrscher, and Kurt Rothermel. On the Impact of Radio Propagation Models on MANET Simulation Results. In *Proceedings of the 7th IFIP International Conference on Mobile and Wireless Communication Networks (MWCN 2005), Marrakech, Morocco, September 2005*, 2005. (Cited on page 20)

[SHR08] Minsu Shin, Seongik Hong, and Injong Rhee. Dtn routing strategies using optimal search patterns. In *CHANTS '08: Proceedings of the third ACM workshop on Challenged networks*, pages 27–32, 2008. (Cited on page 14)

[SNT08] Gabriel Sandulescu and Simin Nadjm-Tehrani. Opportunistic dtn routing with window-aware adaptive replication. In *AINTEC '08: Proceedings of the 4th Asian Conference on Internet Engineering*, pages 103–112, 2008. (Cited on page 14)

[SR06] Illya Stepanov and Kurt Rothermel. On the impact of a more realistic physical layer on MANET simulations results. *Ad Hoc Networks. To appear.*, pages 1–1, 2006. (Cited on page 20)

[SSX07] Bageshree Shevade, Hari Sundaram, and Lexing Xie. Modeling personal and social network context for event annotation in images. In *JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 127–134, 2007. (Cited on page 6)

[TG05] C. Tuduce and T. Gross. A mobility model based on wlan traces and its validation. In *Proceedings IEEE INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, 2005. (Cited on pages 21 and 27)

[TGK06] M. Thomas, A. Gupta, and S. Keshav. Group based routing in disconnected ad hoc networks. *Lecture Notes in Computer Science*, 4297:399–410, 2006. (Cited on page 29)

[TOS+06] R. Toivonen, J.P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski. A model for social networks. *Physica A: Statistical Mechanics and its Applications*, 371(2):851–860, 2006. (Cited on pages vii, 13 and 14)

[Uni] University of Cambridge. Social network founded mobility models for ad hoc network research. `http://www.cl.cam.ac.uk/research/srg/netos/mobilitymodels/index.html`. (Cited on page 73)

[Wat99] D.J. Watts. *Small worlds: the dynamics of networks between order and randomness*. Princeton University Press, 1999. (Cited on page 11)

[WLZ08] L. Wan, J. Liao, and X. Zhu. CDPM: Finding and Evaluating Community Structure in Social Networks. In *Proceedings of the 4th international conference on Advanced Data Mining and Applications*, pages 620–627, 2008. (Cited on page 10)

[WS98] D.J. Watts and S.H. Strogatz. Collective dynamics ofsmall-world'networks. *Nature*, 393(6684):440–442, 1998. (Cited on page 9)

[YHCC07] E. Yoneki, P. Hui, S.Y. 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*, pages 225–234, 2007. (Cited on page 15)

[ZHL06] Qunwei Zheng, Xiaoyan Hong, and Jun Liu. An agenda based mobility model. In *ANSS '06: Proceedings of the 39th annual Symposium on Simulation*, pages 188–195, 2006. (Cited on pages 3 and 21)

[ZSHD05] W. X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar. Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561):439–444, 2005. (Cited on page 10)

All links were last followed on June 11, 2009.

List of Symbols

Symbol	Description	Ref.
$A_{loc}(v, a, t)$	Location attraction exerted by an anchor a towards a node v at time t	48
$A^{I}_{node}(v, a, t)$	Node attraction exerted by an anchor a towards a node v at time t (simple calculation)	49
$A^{II}_{node}(v, a, t)$	Node attraction exerted by an anchor a towards a node v at time t (enhanced calculation)	50
$A_{overall}(G, a, t)$	Overall attraction an anchor a exerts towards a group of nodes $G \subseteq V$ at time t	54
$A_{rep}(v, a, t)$	Node repulsion exerted by an anchor a towards a node v at time t	53
$\mathcal{A}(v)$	Home anchor association function that maps $v \in V$ to its set of home anchors	42
α	Weight between the node and location attraction	54
$c(u, v, t)$	Correction factor of the social relation $\{u, v\} \in E$ at time $t \in T$	59
$C(G)$	Clustering coefficient of a graph G	7
$C_{active}(v, t)$	Active social context of a node v at time t	43
$d(v, u)$	Distance (shortest path) in the social network between two nodes $u, v \in V$	6
D_a	Dwell time distribution of anchor $a \in A$	44
E	Set of social relations in a social network	5
f_a	Anchor function of anchor $a \in A$	46
ϕ	Penalty factor	54
g_{thresh}	Group movement threshold	45
g	Greediness of the destination selection algorithm	54
$L(G)$	Average path length of a graph G	7
$\mathcal{L}(a, t)$	Set of nodes associated with an anchor $a \in A$ at time $t \in T$	44

$m(v,u,t)$	Meetings between v and u until time t	62
$M(v,t)$	Total number of meetings a node $v \in V$ has produced with its active social acquaintances at time t	62
$p_{iso}(v,t)$	Probability that a node $v \in V$ enters into an isolation phase if its current dwell time expires at time $t \in T$	58
$p_{join}(v,a,t)$	Probability that a node $v \in V$ at $a \in A$ joins a group movement at time $t \in T$	45
$remDwellTime(v,a,t)$	Remaining dwell time of node v at anchor a at time t	45
$s(u,v,t)$	Social attraction between two nodes $u, v \in V$ at time t	59
$soc(v,t)$	Sociability of a node $v \in V$ at time $t \in T$	57
T	Set of simulation ticks	43
t_{max}	Length of the simulation	43
t_p	Length of the simulation period	43
t_{update}	Correction factor update interval	62
V	Set of nodes in a social network	5
$w(u,v)$	Weight of the social relation $\{u,v\} \in E$	5
w_{weak}	Penalty threshold	53
$\chi(u,v)$	Set of simulation ticks at which the social relation $\{u,v\} \in E$ is active	43

Erklärung

Hiermit versichere ich, diese Arbeit selbständig verfasst und nur die angegebenen Quellen benutzt zu haben.

(Daniel Fischer)