# Computational Modeling of Spatio-temporal Social Networks: A Time-Aggregated Graph Approach

SHASHI SHEKHAR AND DEV OLIVER
Department of Computer Science & Engineering
University of Minnesota, Minneapolis
**Email:** shekhar@cs.umn.edu; **Web:** http://www.cs.umn.edu/~shekhar

## Introduction

Social computing is transforming on-line spaces with popular applications such as social-networking (e.g., Facebook), collaborative authoring (e.g., Wikipedia), social bargain-hunting (e.g., Groupon), etc. Spatio-temporal constraints are becoming a critical issue in social computing with the emergence of location-based social-networking, Volunteered Geographic Information (Goo 07, Elw 08), Participative Planning (Elw 08, Fis 01), etc. Location-based social networks (e.g., foursquare.com and the "Places Check-in" feature on Facebook) facilitate socialization with nearby friends at restaurants, bars, museums, and concerts. Volunteered Geographic Information (e.g., Wikimapia, OpenStreetMap, Google MyMaps) allows Internet users to participate in generation of geographic information. Traditional computational models for social networks are based on graphs [Fre 06, Was 94, Nrc 03, Cro 09], where nodes represent individual actors (e.g., persons, organizations) and edges represent relationship ties (e.g., communication, financial aid, contracts) between actors. Such graph models are used to assess centrality and the influence of actors (e.g., measures such as degree, reach, "between-ness," bridge), as well as community structure (e.g., measures such as cohesion, clustering, etc.). Statistical properties such as skewed degree distribution are modeled by random graphs [New 02, Nrc 03], where each node-pair has a connecting edge with independent probability $p$, which may depend on factors such as geographic distance [Won 05].

However, traditional graph and random graph models are limited in addressing spatio-temporal questions such as change (e.g., how is trust or leadership changing over time? who are the emerging leaders in a group? what are the recurring changes in a group?), trends (e.g., what are the long-term and short-term trends in network size or structure? what are the exceptions to the long-term trend?), duration (e.g., how long is the tenure of a leader in a group? how long does it take to elevate the level of trust such as a relationship changing from visitor to friend?), migration, mobility and travel (e.g., interplay between travel behavior and size/structure of social networks [Tim 06]). This position paper explores time-aggregated graph models to support computational tools to address such questions.

## Background

"A social network is a social structure made up of individuals (or organizations) called "nodes," which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige" [Wik 10]. Network formation theories bring

up the principle of homophily [McP 01] (i.e. birds of a feather flock together) and differentiate between two types, namely, baseline and in-breeding. Baseline homophily [Fel 81] refers to the limited social pool available to individuals for tie formation due to foci of activities, demographics, etc. In-breeding homophily models additional constraints due to gender, religion, social class, education, personality, etc.

Spatio-temporal constraints (e.g., geographic space, travel, schedules and diurnal cycles) play a major role in determining baseline homophily due to reasons like opportunity and minimization of cost and effort [Deb 69]. Survey research on student housing communities [Ath 73, Mok 04] and Torontonian personal communities [Wel 88] have provided evidence on the role of geographic spaces. For example, [Wel 88] noted that about 42% of "frequent contact" ties lived within a mile of a typical person. Computational simulation based on agent-based models [Cro 09], game-theory and cost-benefit analysis [Joh 00] as well as spatially embedded random graphs and distance-decay based edge probabilities [Won 05] have reproduced many properties of social networks including small-world properties (e.g., graph diameter), short average geographic distances, community structures, low tie density, etc.

Traditionally, social network research has had a relatively small amount of data from infrequent longitudinal surveys and computer simulations. However, the recent popularity of social computing on the Internet is providing large and frequently-sampled spatio-temporal social interaction datasets. These may facilitate better understanding of social network formation and the role that spatio-temporal constraints play, in the face of opportunities to interact with distant actors via Internet based social networking applications. This development also highlights a central role for computation and computational models, not only to scale up to the large and growing data volumes, but also to address new spatio-temporal social questions related to change, trends, duration, mobility, and travel.

## Time-Aggregated Graphs

Given a spatio-temporal social network dataset and analysis questions, a computational model provides a representation to not only specify the dataset and questions, but also design data-structures and algorithms for addressing the questions. The model should be able to scale as large datasets are becoming available from Internet based social computing services with a large number of actors and a large number of time-points.

There are several challenges associated with modeling such a network. The model should be able to accommodate changes and compute results consistent with the existing conditions both accurately and simply. Furthermore, for quickly answering frequent queries such as tracking recurring changes in a group, fast algorithms are required for computing the query results. Sufficient support for the design of correct and efficient algorithms should therefore be provided by the model. The need for computational efficiency conflicts with the requirement for expressive power of the model and balancing these two conflicting goals is challenging.
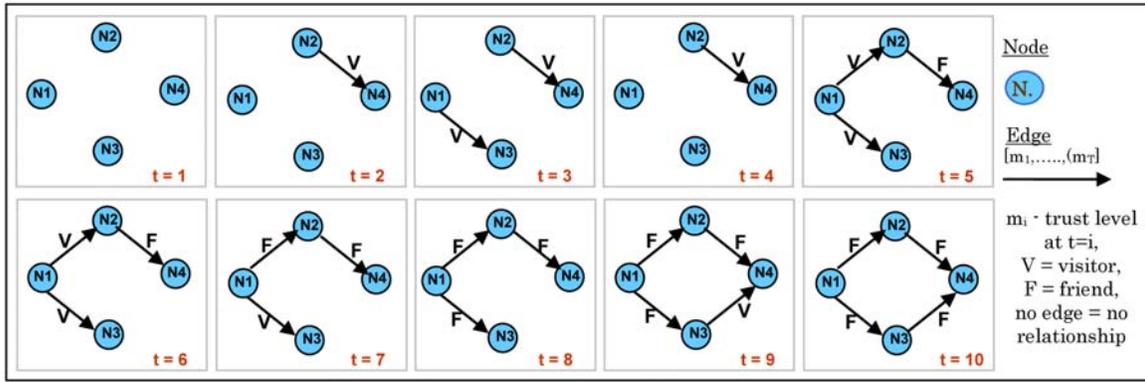
*Figure 1: A Time-series of Snapshots for a trust Network for time instants 1 through 10*

Dynamic networks are often modeled using a time-series of snapshots [Tan 07, Lah 08] of actors and their relationships. For example, Figure 1 shows a time-series of snapshots of a trust network for time instants, t = 1 to t = 10. Three levels of trust are exhibited in Figure 1, where an absent edge indicates that there is no trust relationship, an edge with V indicates a visitor trust relationship and an edge with F indicates friendship, which is a stronger relationship than visitor. For example, (N1, N3) have no trust relationship at t = 1, 2, and 4, a visitor relationship at t = 3, 5, 6, and 7 and a friend relationship at t = 8, 9, and 10. Due to duplication of information about nodes and edges across snapshots, the scalability of snapshot time-series model is limited in answering longitudinal questions such as how long it takes a relationship to evolve from visitor to friend. Storage and thus computational cost increases linearly with the number of time-points.

An alternative is the time aggregated graph (TAG) model [Geo 06], which has provided scalable algorithms for temporal questions in transportation networks [Geo 08]. Figure 2 shows a time aggregated graph that is based on the trust network of Figure 1. Each edge has a time series (enclosed in square brackets). For example, the trust relationship for the edge (N1, N3) for all instants within the time interval under consideration are aggregated into a time series [-,-,V,-,V,V,V,F,F,F]; the entry '-' indicates that the edge is absent at the time instants t = 1, t = 2 and t = 4.
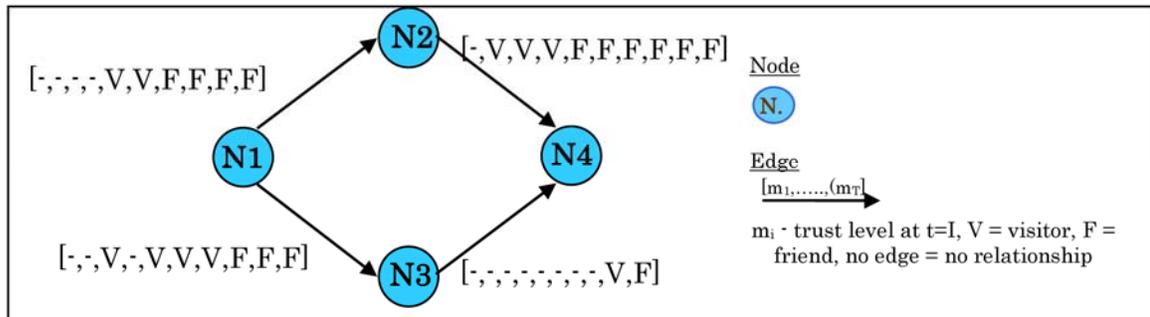


*Figure 2: A Time Aggregated Graph for a Trust Network for time instants 1 through 10*

With the time aggregated graph model, the longitudinal behavior is captured for each edge (or node). Time aggregated graphs do not replicate information about nodes and edges across time-points. This reduces storage cost as well as computational costs. It also consolidates time-series information to make it easier to answer cross snapshot questions such as how long it takes for a relationship to evolve from visitor to friend. For example, in Figure 3, it takes 2 units for edge (N1, N2), 3 units for edge (N2, N4), 4 units for edge (N1, N3) and 1 unit for edge (N3, N4) with an average of (2 + 3 + 4 + 1) / 4 = 2 units, for a relationship to evolve from visitor to friend. Answering this duration question in snapshot model takes more effort due to the need to repeatedly go from one snapshot to the other.

## Discussion

Time-Aggregated Graphs has the potential to be a general representation of temporal evolution of relationships, whose snapshots were traditionally modeled as graphs. Thus, it may help address questions about individual relationships over longer time-frames by providing a representation for relevant datasets. For example, consider social network datasets representing *friend-of* relationships among people. Currently, graph representations are used to model a static (e.g., time-snapshot) view of social relationships to explore questions like centrality (e.g., leaders), and cohesiveness of communities. In contrast, TAG may support a direct representation of a "friend-of" relationship over long time periods to address questions related to changes in and evolution of centrality (e.g., emerging leaders) and group cohesiveness (e.g., increasing, diminishing). We welcome collaboration towards identifying datasets and use-cases to evaluate the potential of TAG to address spatio-temporal questions about social networks.

## References:

[And 08] C. F. d'Andrea, Wikipedia as social interaction rooms and collaborative-writing processes, WikiSym, ACM 978-1-60558-128, 2008.

[Ath 73] R. Athanasiou, G. A. Yoshioka, The spatial characteristics of friendship formation, *Environ. Behav*, 5, 1973, 43–66.

[Cro 09] Crossman, J., Bechtel, R., Parunak, H., and Brueckner, S., Integrating dynamic social networks and spatio-temporal models for risk assessment, wargaming and planning. In *The Network Science Workshop*. 2009, West Point, NY.

[Deb 69] G. Debreu, Neighboring economic agents, *La Decision*, 171: 85–90, 1969.

[Elw 08] S. Elwood, Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS. *GeoJournal* 72(3 & 4): 173–183, 2008.

[Fel 81] S. L. Feld, The focused organization of social ties, *Am. J. Soc.*, 86, 1981, 1015–1035.

[Fis 01] F. Fisher, Building Bridges through Participatory Planning, UN-Habitat, isbn 92-1-131623-5, 2001. www.gdrc.org/decision/BuildingBridges.pdf.

[Fre 06] L. Freeman, *The Development of Social Network Analysis*. Vancouver: Empirical Pres, 2006.

[Geo 08] B. George, S. Shekhar, and S. Kim. Spatio-temporal network databases and routing algorithms. University of Minnesota. CSE Technical Report No. 08-039, 2008. (A summary of results in Proc. Symposium on Spatial and Temporal Databases, Springer LNCS 4605, 2007).

[Geo 06] B. George, S. Shekhar, Time-Aggregated Graphs for Modeling Spatio-temporal Networks, *Journal of Data Semantics*, 11, 2006 (Springer LNCS 5383, 2008). (Special issue on best papers from ER 2006 Workshop on Conceptual Modeling for GIS).

[Goo 07] M. F. Goodchild, Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4): 211–221, 2007.

[Joh 00] C. Johnson and R. P. Giles, *Spatial Social Networks, Review of Economic Design*, 5, 273–299, Springer Verlag, 2000.

[Lah 08] M. Lahiri, T. Y. Berger-Wolf, Mining Periodic Behavior in Dynamic Social Networks, Proc. ICDM, IEEE, 2008.

[McP 01] M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks, *Annu. Rev. Sociol.*, 27, 2001, 415–444.

[Mok 04] D. Mok, B. Wellman, R. Basu, Does distance matter for relationships? SUNBELT International Social Network Conference, Portoroz, Slovenia, 2004.

[New 02] M. E. Newman, D. J. Watts, S. Strogatz, Random graph models of social networks, Proceedings of the National Academy of Science, 99, 2566–2572, Feb. 19th, 2002.

[Nrc 03] National Research Council, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, isbn 0-309-08952-2, National Academies Press, 2003.

[Tan 07] C. Tantipathananandh, T. Berger-Wolf, D. Kempe, A framework for community identification in dynamic social networks, Proc. SIGKDD, ACM, 2007.

[Tap 06] D. Tapscott, A. Willians. *Wikiconomics: How mass collaboration changes everything.* New York: Portifolio, 2006.

[Tim 06] O. Timo, Mapping Social Networks in Time and Space, Arbeitsberichte Verkehr und Raumplanung, Working Paper 341, IVT, ETH Zurich, 2006.

[Was 94] S. Wasserman, and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.

[Wel 88] B. Wellman, P. Carrington, A. Hall, Networks as personal communities, in B. Wellman, S. D. Berkowitz (Eds.), *Social Structures: A Network Approach*, Cambridge University Press, Cambridge, 1988.

[Wik 10] Social Network, Wikipedia, http://en.wikipedia.org/wiki/Social_network, retrieved Nov. 29, 2010.

[Won 06] L. H. Wong, P. Pattison, G. Robbins, A spatial model for social networks, *Physica A*, 360, 2006, 99–120, Elsevier.