

# Space, Time, and Sociability: Predicting Future Interactions

Christpoh Stich<sup>\*1</sup>

<sup>1</sup>School of Geography, Earth, and Environmental Sciences, University of  
Birmingham

January 13, 2017

## Summary

While space, time and the social realm are intrinsically linked hardly any studies have taken their effects into account at the same time. We use data collected by the Copenhagen network study that tracked 847 students via mobile phones. We assess for the first the time the predictive power of spatial, temporal, and social variables simultaneously on predicting physical interactions. We find that social variables and to a lesser extent temporal variables have the greatest impact on prediction performance. It almost seems as if the only thing that matters are the people you meet and not where you meet them.

**KEYWORDS:** *geographic context, link-prediction, social networks*

---

<sup>\*</sup>cxs489@student.bham.ac.uk

## Introduction

Space, time and the social realm are intrinsically linked. For the most part people do not aimlessly amble, nor is happenstance usually the reason people spend time at places. People commute to work every day, they meet their friends at a bar after work, or they go on a date with their partner. In fact, Eagle and Pentland (2009) by employing principal component analysis and Song et al. (2010) by looking at mobile entropy find that people are highly predictable in their daily movement patterns.

While an extensive amount of research has already been conducted on the interplay between the social realm, place, and time, studies so far were either limited to a very specific type of network or did not jointly deal with all three factors. On the one hand, several studies that have accounted for spatial and temporal features have focused on a narrow set of social interactions such as online social networks or face-to-face networks. One group of research projects studied very topical online social networks such as the Foursquare network (Scellato et al., 2011) or the Flickr (Crandall et al., 2010) network, while another group have focused on studies of face-to-face interactions solely in highly structured and defined settings—a museum, a conference, and a primary school (Isella et al., 2011; Stehlé et al., 2011; Zhao et al., 2011). Last but not least, Noulas et al (2015) analyze spatial, temporal, and social features but focus on networks of places instead of individuals.

On the other hand studies that looked at more broadly defined social networks have not assessed spatial and temporal features at the same time. Although Yang et al (2013) uses information about when and in which network configuration people have met as features for their link-prediction algorithm, they do not incorporate spatial features. Sekara et al (2015) construct social cores around which interactions are structured based on social and temporal variables. However, place does not play a role in their subsequent prediction task. While, Wang et al (2011) successfully uses the similarity of trajectories of users for predicting phone calls between users, they do not take any other temporal or spatial features into account.

We believe *a joint assessment of spatial, temporal, and social features to be crucial for understanding the true dynamics behind social interactions* as human interactions might be spatially, temporally, and/or socially confounded with each other.

Consequently, our contribution consists of two parts: a) ascertaining whether geographic places themselves hold discriminatory power or are merely a container for interaction b) assessing the “simultaneous” predictive information of geographic, temporal, and social features for an ever changing network of interactions. In short, we try to better understand what factors drive the evolution of a human social interaction network, and how we might use salient features for predicting future interactions.

To address those two questions, we phrase the question of whether two people will meet in

a given time period as a link-prediction problem in a social interaction graph. We train a random forest classifier (Breiman, 2001) on various subsets of our feature space to assess the predictive power of time, place, and social features for predicting future links in the graph. We utilize data from the Copenhagen network study to run our experiments. This dataset is unique as it captures GPS, Bluetooth and WiFi signals, calls and texts for a relatively large group of 847 students over the course of a whole year. And it is especially suited for our problem as it provides us with a detailed view of everyday life and interactions among “normal” peers.

## Link Prediction

Random forests have been consistently shown to perform well in link-prediction tasks (Peng et al., 2015) and we thus opt to use them for our prediction task as well (Pedregosa and Varoquaux, 2011).

## Feature Vectors

### Baseline Features

As our baseline features for all subsequent models we include *recency*, the amount of elapsed time since the last meeting (Yang et al., 2013), *activeness*, how often two nodes interacted (Yang et al., 2013), and the amount of time we observe two nodes together.

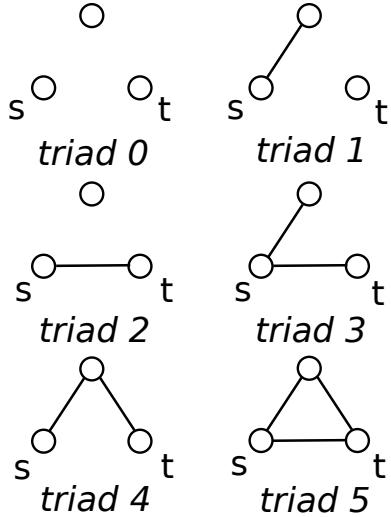
### Spatial, Temporal, and Social Features

We also include several features pertaining to the setting wherein two nodes meet. These can be split into features relating to time, space, and the social realm. The time related features pertain to capture weekly behavioral patterns. Let  $M$  be the set of all meetings between two nodes  $u, v \in G$  in the training period. We then include a  $\text{vector}((\text{hour-of-week}(M))$  that counts the interaction between  $u$  and  $v$  at every hour of the week.

We also include  $\min(\text{place entropy}(u, v))$  (Scellato et al., 2011) of the meetings between two students  $u$  and  $v$ . We also infer the *relative importance* of each venue for each user  $u$  by measuring the amount of time a user spends there and ranking the venues by the amount of time spent there. We then include the  $\maxRank(\text{relative importance}(u, v))$  of a meeting between any  $u, v \in G$ .

Let  $\text{context}(M_{u,v})$  be the function that counts the amount of time two nodes  $u, v \in G$  have spent together at the different geographic contexts (either home, university, an im-

Figure 1: Triadic periods



The figure shows all possible combination of co-locations of any three students in our interaction graph.

portant other place where a student spends a lot o time or other). We then include the *context*( $M_{u,v}$ ) as a feature as well.

Let  $P_{u,v}$  be the distribution of the number of other people from the study that are present when two nodes  $u, v \in G$  meet. We then include  $\text{avg}(P_{u,v})$  as a feature.

Triadic-closure, the phenomenon in social network that friends of friends are likely to become friends themselves, has been known to play a significant role in network formation (Bianconi et al., 2014). Yang et al. (2013) have used the “triadic periods” successfully as features for predicting interactions before. We build upon their work and adapt their metric for our problem. The main idea is to count the different possible arrangements of triads in the interaction graph, or in other words the different possible configurations of co-locations. Figure 1 shows the possible arrangements of co-location triads (excluding symmetric triads).

## Network Features

We also include several features that are based on the network topology of the interaction graph and have been used extensively in link-prediction problems before. In particular we include Adamic Adar coefficient (AA), the Jaccard coefficient (JC), preferential attachment (PA), ressource allocation (RA), prop flow (PF) and weighted prop flow (weighted PF) (Peng et al., 2015).

## **Node Features**

Sometimes one however might not have access to the whole network and one might only be in possession of more or less isolated node level data. We simulate such a scenario by building another model that only incorporates features that can be obtained from the ego-network of a node. In particular the features for the node model are: The baseline features, and all the spatial, temporal, and social features with the limitation that “triad 4” cannot be distinguished from “triad 1” and “triad 5” not from “triad 3”.

## **Model Construction**

Besides building null model based on random changes of ties between time periods, we construct a *base* model that only contains the baseline features. We build the *node*, *time*, *social*, *place*, and *network* models by adding to the base models the features that pertain to that domain (as described above). The *context* model consists of the baseline features as well as the temporal, spatial, and social features. The *full* model consists of all features. Last but not least after our experiment we construct a *refactored* model based on top five features of the *full* model.

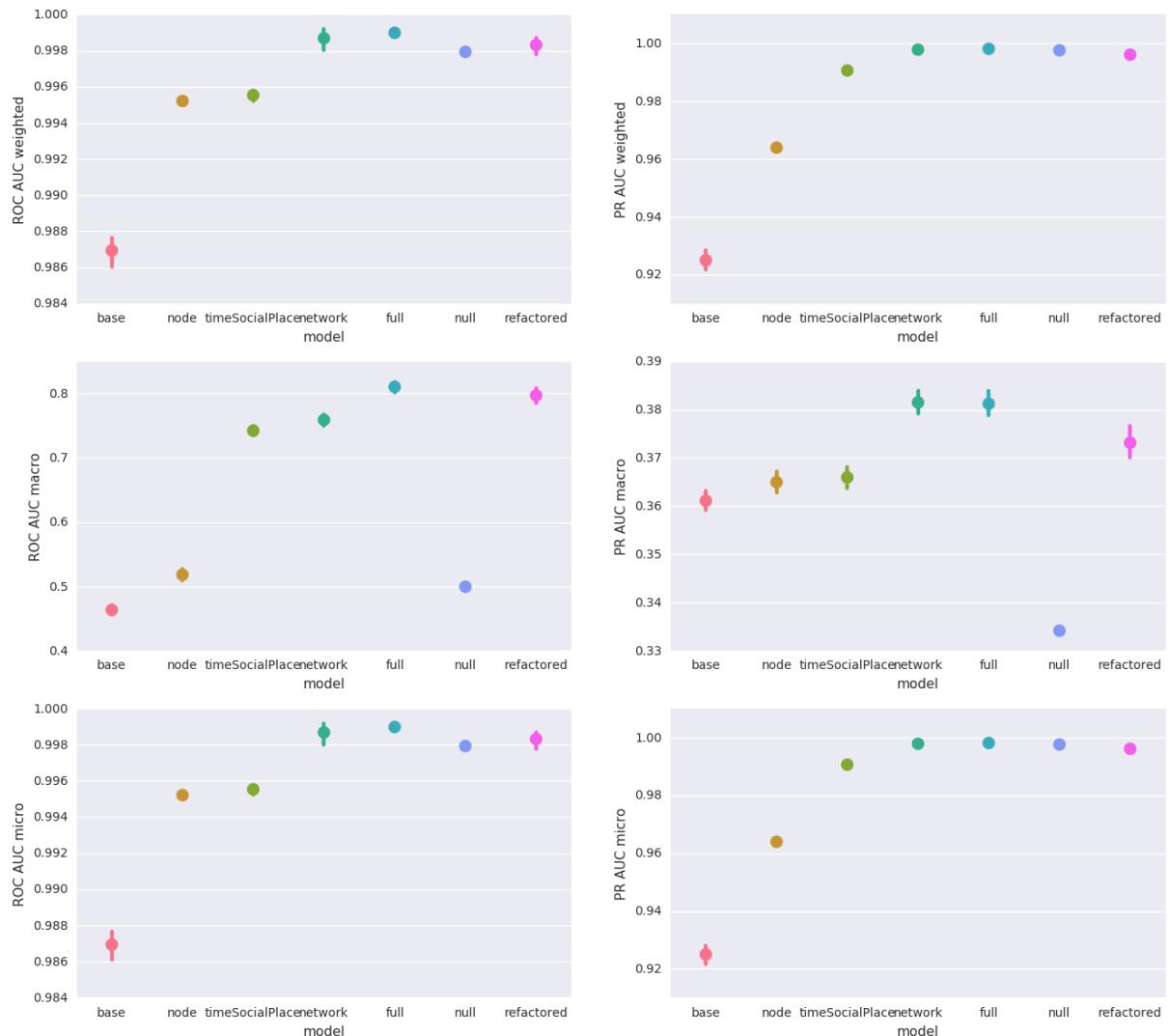
## **Findings**

To compare the performance of our different models we have chosen to report the receiver operating characteristic area-under-the-curve (ROC AUC) as well as precision-recall area-under-the-curve (PR AUC). We have opted to report both metrics as Davis and Goadrich (2006) have shown that ROC curves might provide an overly optimistic evaluation of the algorithm.

## **Performance of the Link-prediction Algorithm**

First of all, our full model performs either significantly better or as least as good as the null model (see Figure 2). The high performance of the null model as compared to the other models when *micro* averaging our classes can explained by the large class imbalance in our data. Class 0 (meaning that two nodes are not meeting) is overly the dominant. Second, the difference between the performance of the network model and the full model is small. Third, given the relatively low performance of the *base*, *node*, and *context* models compared to the *network* and *full* model, one can conclude that features derived from the interaction network are essential for predicting future interactions between nodes. Consequently, the

Figure 2: Model scores



The different area-under-the-curve scores for our models.

refactored model that includes several network measures performs almost as well as the full model.

## Conclusion

We find that *when* and *whom* you meet holds the most discriminatory power for predicting links, while *where* you have met has a negligible impact on prediction. While we do not see an a priori reason why our findings among students should not hold for the general population, we have to acknowledge that we consider only a bias and limited sample for our study. At the same time, however, it is noteworthy that the Copenhagen network study covers a significant portion of the freshman of the Technical University of Denmark.

## 1 Acknowledgements

The research is part of my doctoral thesis at the University of Birmingham. My research is funded in part by the ESRC as well as the School of Geography, Earth, and Environmental Sciences. I am grateful for the continued support of my supervisors Emmanouil Tranos, University of Birmingham, and Mirco Musolesi, UCL. Last but not least I am grateful to Sune Lehman, DTU, for providing access to the data as well as further support.

## 2 Biography

I am a PhD research student at the University of Birmingham and I just finished my second year. My research interests lie in the intersection of human behavior, geography, and “Big Data”. I am particularly interested in how human mobility shapes social interactions and vice versa.

## References

- Bianconi, G., Darst, R. K., Iacovacci, J., and Fortunato, S. (2014). Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E*, 90(4):20–29. (document)
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32. (document)
- Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., and Kleinberg, J. (2010). Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52):22436–41. (document)

- Davis, J. and Goadrich, M. (2006). Relationship between Precision-Recall and ROC curves. *Proceedings of ICML'06*. (document)
- Eagle, N. and Pentland, A. S. (2009). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066. (document)
- Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., and Van den Broeck, W. (2011). What’s in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–80. (document)
- Noulas, A., Shaw, B., Lambiotte, R., and Mascolo, C. (2015). Topological Properties and Temporal Dynamics of Place Networks in Urban Environments. *WWW ’15 Companion Proceedings of the 24th International Conference on World Wide Web*, pages 431–441. (document)
- Pedregosa, F. and Varoquaux, G. (2011). Scikit-learn: Machine learning in Python. . . of *Machine Learning* . . . , 12:2825–2830. (document)
- Peng, W., Baowen, X. U., Yurong, W. U., and Xiaoyu, Z. (2015). Link Prediction in Social Networks : the State-of-the-Art. 58(January):1–38. (document)
- Scellato, S., Noulas, A., and Mascolo, C. (2011). Exploiting Place Features in Link Prediction on Location-based Social Networks Categories and Subject Descriptors. *Proceedings of KDD’11*, (Section 3):1046–1054. (document)
- Sekara, V., Stopczynski, A., and Lehmann, S. (2015). The fundamental structures of dynamic social networks. *arXiv preprint arXiv:1506.04704*, page 10. (document)
- Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science (New York, N.Y.)*, 327(5968):1018–21. (document)
- Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B., and Vanhems, P. (2011). High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLoS ONE*, 6(8):e23176. (document)
- Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.-l., and Science, C. (2011). Human Mobility , Social Ties , and Link Prediction Categories and Subject Descriptors. *Proceedings of KDD’11*. (document)
- Yang, Y., Chawla, N., Basu, P., Prabhala, B., and La Porta, T. (2013). Link prediction in human mobility networks. *Proceedings of ASONAM 2013*, 1(c):380–387. (document)
- Zhao, K., Stehlé, J., Bianconi, G., and Barrat, A. (2011). Social network dynamics of face-to-face interactions. *Physical Review E*, 83(5):056109. (document)