Human Contact Prediction Using Contact Graph Inference

Kazem Jahanbakhsh, Gholamali C. Shoja, and Valerie King
Email: jahan@cs.uvic.ca

Computer Science Department, University of Victoria

Dec 20, 2010

Outline I

Outline II

Outline III

Human Mobility Prediction

- Predicting how people move is a complex problem.
- Applications:
  - Traffic planning in cities.
  - Modeling epidemic disease.
  - Application for network communication.
- We formulate the problem of human contact prediction as a graph inference problem.

Definitions

- Two people are *in contact* if they are in each other's proximity ($< 10m$).
- Contact events among people can be shown by a weighted graph called *contact graph*.
- In contact graph $G = (V, E)$, V is the set of people and E is the set of contacts among them.
- The weight of each edge (u, v) is the total time they have spent with each other.

Problem Definition I

- The main problem is graph inference when a prior knowledge about the graph is available.
- For the first part of our analysis, we assume that we only have social profiles of nodes in V.
- We assume all edges of G are unknown.
- The problem is to infer edges in E by using the available social information of nodes.

Problem Definition II

- For the second part, we assume that $V = V_{int} \cup V_{ext}$.
- V_{int} and V_{ext} denote the internal and external vertices.
- We assume that all edges in $E_{known} \subset V_{int} \times (V_{int} \cup V_{ext})$ are known.
- However, all edges in $E_{unknown} \subset V_{ext} \times V_{ext}$ are missing.
- Our problem is to infer the edges among external vertices (e.g. $E_{unknown}$).

Related Work

- Eagle et al. and Mitbaa et al. have shown a close relation between people's mobility and their friendship network.
- Daly et al. and Hui et al. proposed routing algorithms which exploit contact graphs properties.
- Nowell and Kleinberg have studied the problem of link prediction in citation networks.
- Goldberg et al. have used cohesive neighborhoods between proteins for assessing the confidence of observed interactions among them.
- Vert et al. have studied the graph inference problem in metabolic networks by employing a supervised learning algorithm.

Real Data Description

- For our analysis, we use the human mobility traces collected from two different conferences.
- The first dataset is collected during the Infocom 2005 conference where 41 participants attended.
- The second dataset is the sampled contacts among 79 people attending Infocom 2006 conference.
- In both experiments, Bluetooth sensors sampled a contact between two people when they were in close proximity of each other (e.g. < 10 meters).
- For Infocom 2006, social profiles of people who participated in the experiment were also collected.

Exploiting Social Information

- Our social profiles contain information about 6 different social dimensions.
- Each social dimension can be shown with a set of social features.
- Suppose node u speaks English and Spanish.
- Let us denote *English* and *Spanish* with 1 and 2 respectively.
- We can show the spoken languages of node u with a feature set $\Gamma u = \{1, 2\}$.

Jacard Social Similarity

- Social similarity between two nodes with respect to dimension i can be computed by using Jacard index:
- $$\sigma^i_{jacard}(u,v) = \frac{|\Gamma^i_u \bigcap \Gamma^i_v|}{|\Gamma^i_u \bigcup \Gamma^i_v|} \quad (1)$$
- Total similarity between two nodes is computed as the average over all dimensions:
- $$sim_{jac}(u,v) = \sum_{i=1}^{d} \frac{\sigma^i_{jacard}(u,v)}{d} \quad (2)$$

Social Foci Similarity

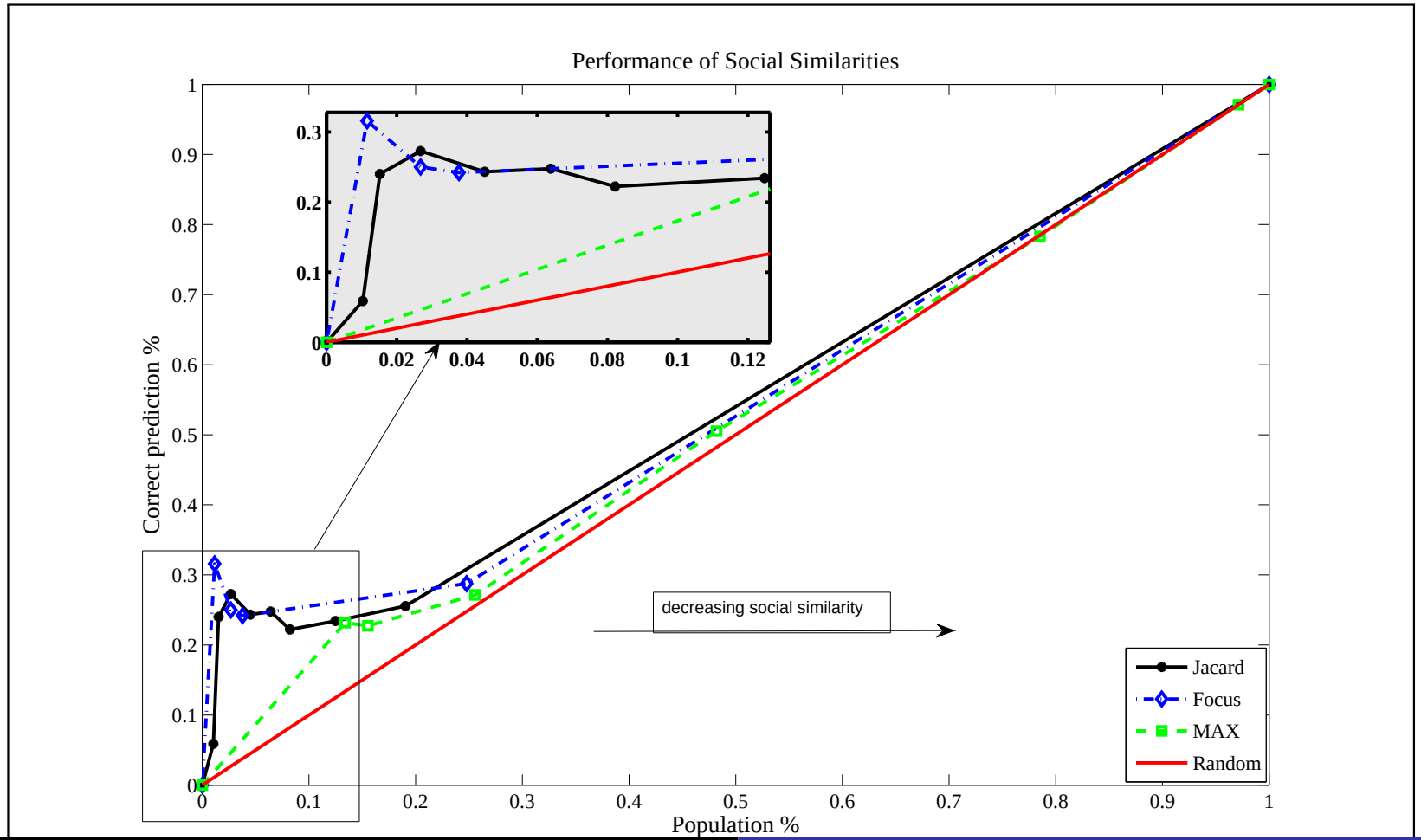
- The social distance between two nodes u and v can be defined as the size of the smallest social feature set that includes both of them:
- $$d_{foc}(u, v) = min\,|\{F|u, v \in F\}| \tag{3}$$
- Here, F is the focus set to which both u and v belong.
- Using the foci distance, the foci similarity between two nodes u and v is:
- $$sim_{foc}(u, v) = \frac{1}{d_{foc}(u, v)} \tag{4}$$

Max Social Similarity

- Finally, the social similarity between two nodes can be defined as the maximum Jacard similarities among all dimensions:
-

$$sim_{max}(u, v) = \max_{i} \sigma^{i}_{jacard}(u, v) \tag{5}$$

Performance of Social Similarities (Infocom 06)



Discussions

- The performances of three social similarities are statistically more significant than random predictor.
- Our results show if two nodes are socially similar, they are more likely to meet.
- Foci distance performs better than the Jacard and MAX.
- As we decrease the similarity threshold, the effect of social profiles diminishes and our results become more similar to random predictor.

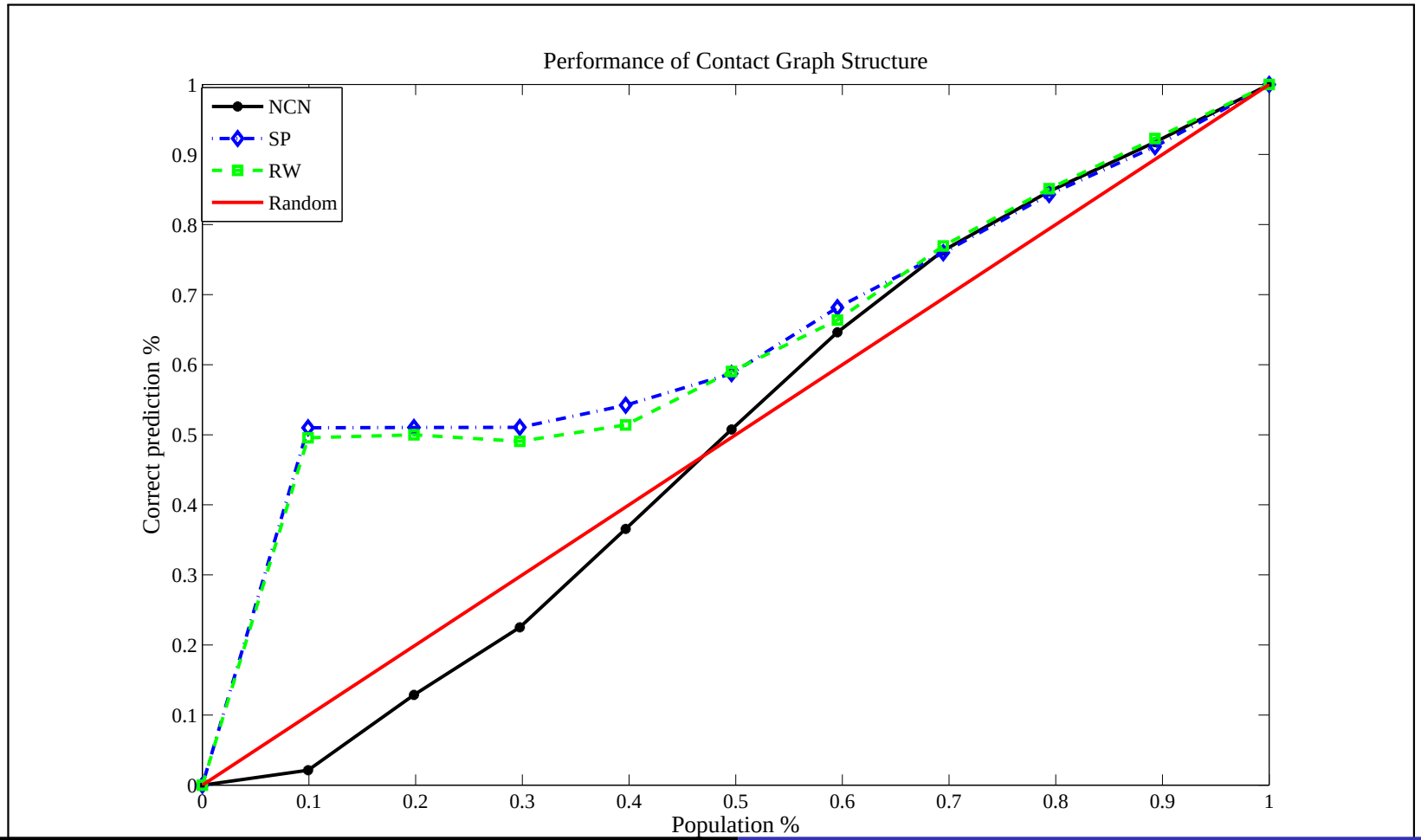
Exploiting Small-World Properties

- *Triangle Closure:* in social networks, if u and v have a common friend w, then there is a high probability for u and v to be friend.
- Social networks have *low diameter*.
- We expect contact graphs also have these properties.
- Now, we want to propose three similarity measures based on the above properties.

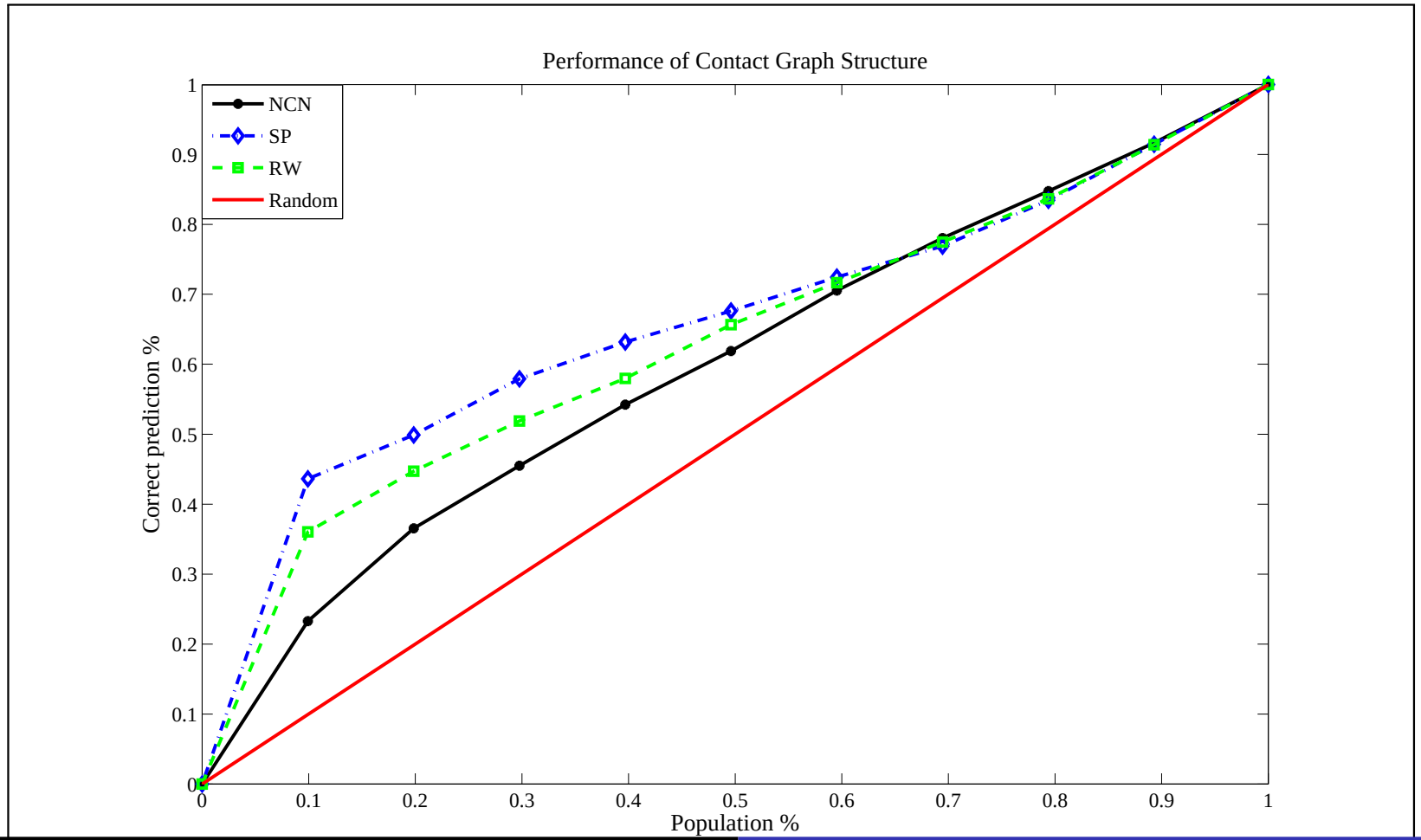
Similarity Measures Based on Small-World Properties

- *Number of Common Neihbors (NCN)*: similarity between u and v can be computed by the number of common neighbors between them.
- *Shortest Path (SP)*: similarity between u and v can be computed by the total weight of the shortest path between them.
- *Random Walk (RW)*: similarity between u and v can be computed by the stationary probability of v for a random walk that starts from u.

Performance of Contact Graph Structure I (Infocom 05)



Performance of Contact Graph Structure II (Infocom 06)



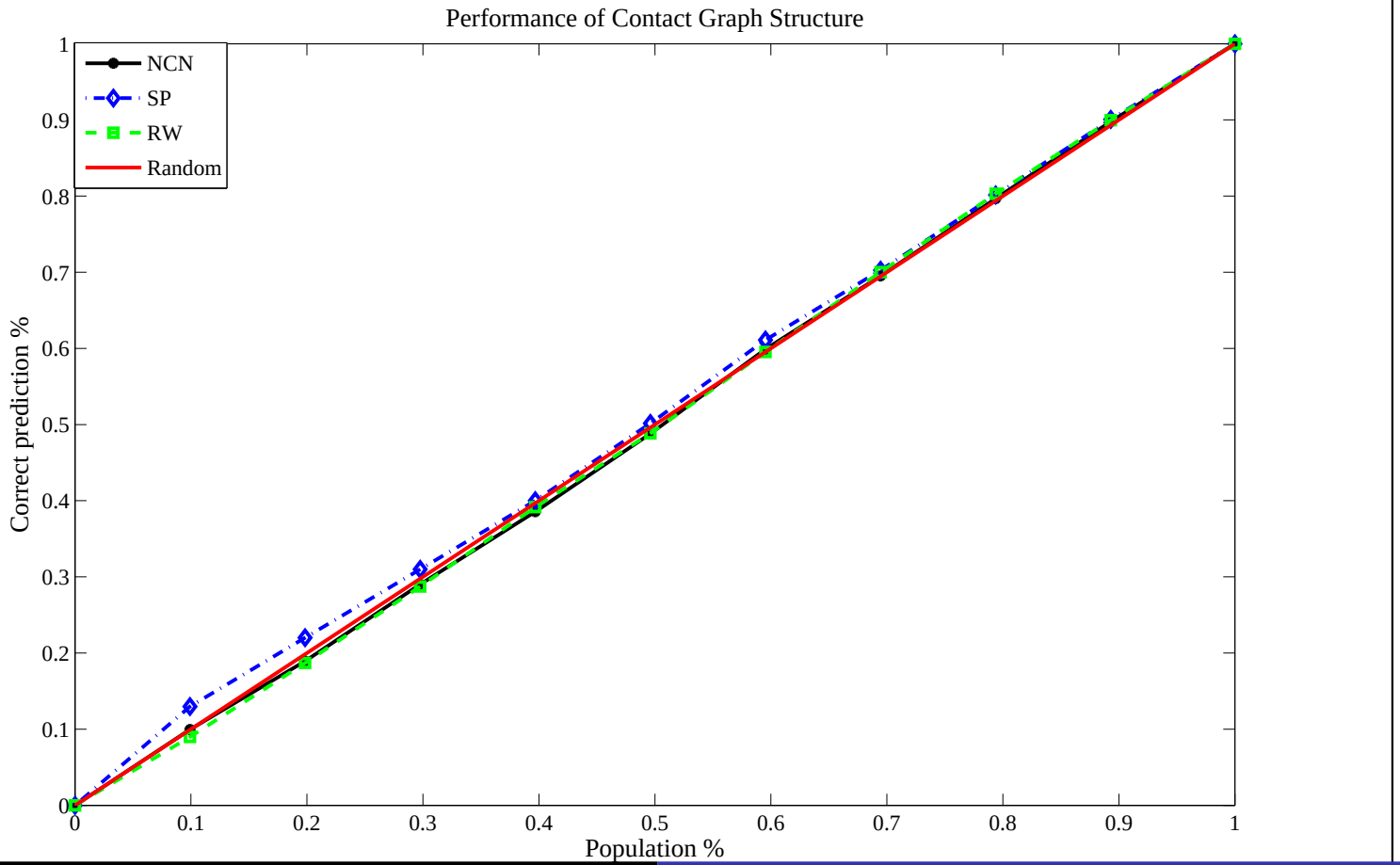
Discussions

- Our results show that both SP and RW outperform the random predictor.
- SP is the best predictor and the RW has the second rank.
- There is an underlying mechanism governing the formation of links in a contact graph that cannot be explained by a purely random process.
- A local maximum happens in the beginning of the graphs where chosen nodes are very similar.

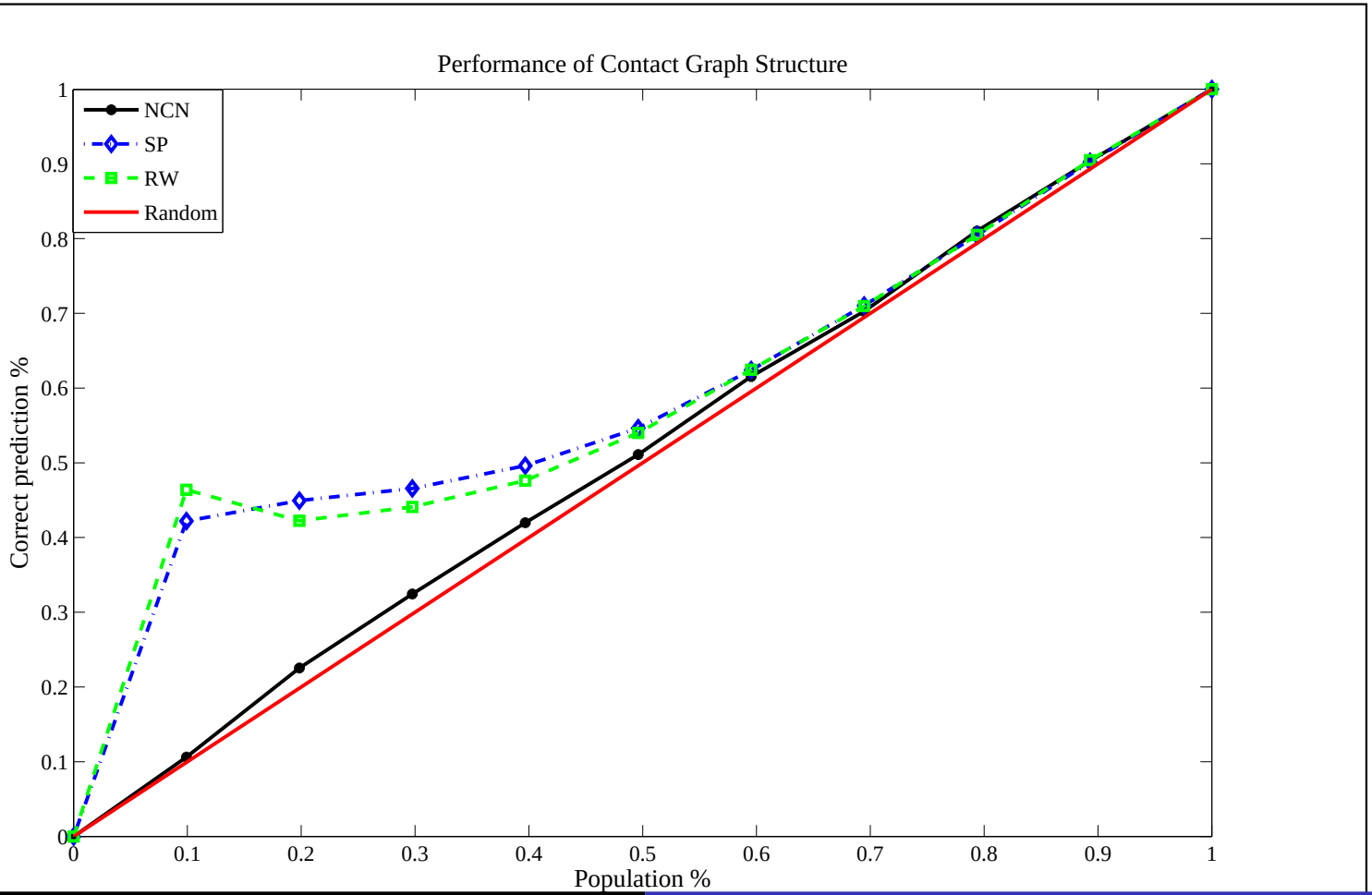
Contact Graph Model

- We use Watts and Strogatz small-world network to model the underlying social graph among a set of nodes.
- The size of graph is $N = 100$.
- We randomly pick a node u from all N nodes.
- The node v which is going to meet u is chosen with a probability q randomly from all N nodes.
- This models the fact that nodes may contact each other at random.
- With probability of $1 - q$, the peer node for u is chosen from one of its friends.
- This models the fact that similar nodes are more likely to see each other.

Performance of Contact Graph Model (I) ($q = 1.0$)



Performance of Contact Graph Model (II) ($q = 0.2$)



Discussions

- When $q = 1.0$, none of predictors performs better than a random predictor.
- In this case, there is not any structure in the simulated contact graph.
- However, for $q = 0.2$ we see similar patterns for SP and RW predictors as the observed ones for real data.
- Thus, homophily process explains the observed local maximums in prediction results.

Conclusions

- We have formulated the problem of contact prediction as a graph inference problem.
- We have shown the importance of using social profiles for contact prediction task.
- The effectiveness of using the underlying properties of contact graphs for contact prediction problem was shown.
- Finally, we have shown the importance of homophily process in the structures of contact graphs.