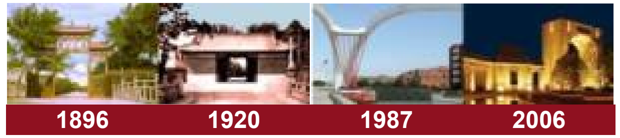
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY



Social Network De-anonymization with Overlapping Communities

5140219173 吴昕宇 大三EE

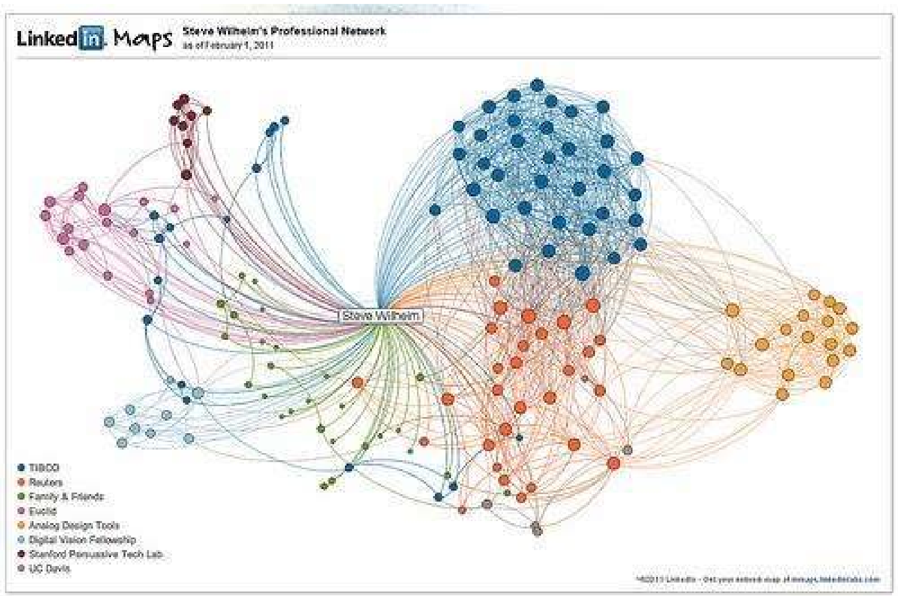
Outline:

1. Problem Background & Formulation

2. Main Contributions

3. Conclusion & Future Work

1. Problem Background & Formulation



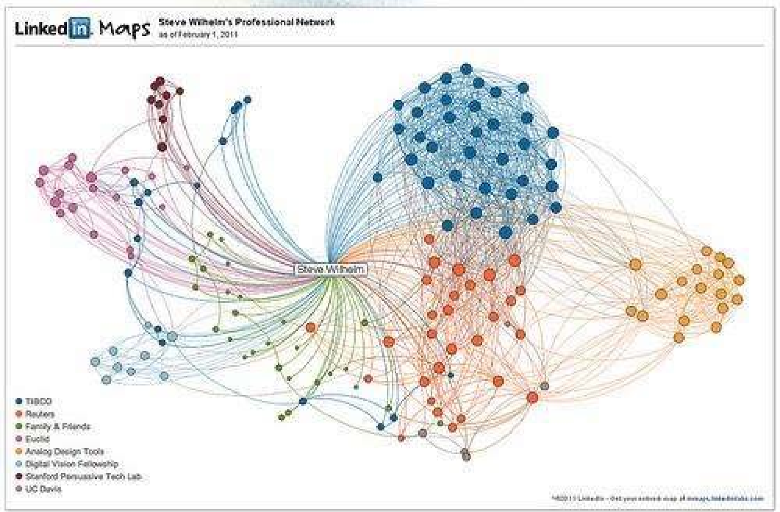
Anonymized Network

Which guy does every node represent?

De-anonymization

1. Problem Background & Formulation

With the prior knowledge of another social network...



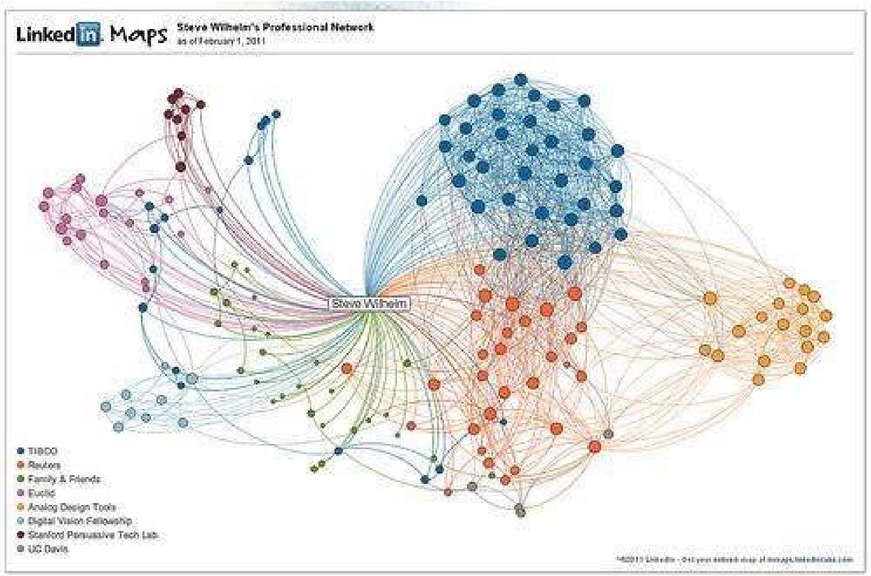
Network A
(Auxiliary Network)

Network B
(Published Network)

Network B: topology & assignment
Network A: topology

} Network A: assignment

1. Problem Background & Formulation



Colors——Communities

Overlapping Communities

Network B: community & topology & assignment
Network A: community & topology
} **Network A: assignment**

1. Problem Background & Formulation

Motivation and Significance of our work:

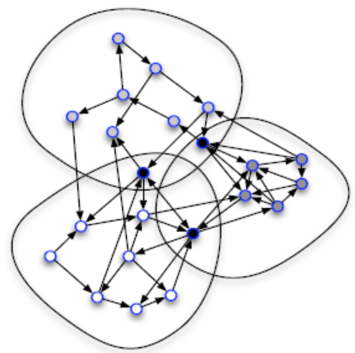
- **Act as a privacy attacker**
- **Study under what conditions they can de-anonymize the network**
- **Protect user privacy according to the study**

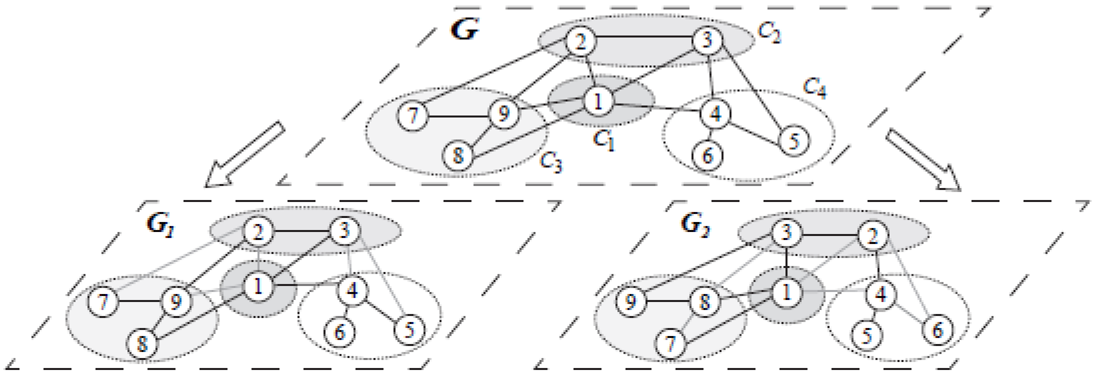
1. Problem Background & Formulation

Mathematical Model:

Social Network ⇔ Graph

- **Users ⇔ Nodes**
- **Relationship ⇔ Edges & Communities**





Overlapping Stochastic Block Model

Correct mapping: 1-1,2-3,3-2,4-4,5-6,6-5,7-9,8-7,9-8

Some definitions:

A,B——Adjacent matrices

M——Community assignment matrix

(M(i,j)=1 means node i is in community j)

π——Permutation matrix

(π(i,j)=1 means node i in Graph B is mapped to node j in Graph A)

2. Main Contributions

Contribution 1:

- **Transform the node matching problem into the edge matching problem.**
- **Prove that the node matching error is bounded by the edge matching error.**
- **Find that the node matching error vanishes compared with the size of the graph when two graphs are similar in degree distribution.**

Node Matching: (Original problem)

minimize $\| \pi - \pi_0 \|$

π **: Our estimation mapping** π_0**: True mapping**

Incentive:

- **No prior information about** π_0**, but much prior info about edges(adjacent matrix)**

Edge Matching: (Transformed problem)

minimize $\|A - \pi B \pi^T\|$ **(+** $\mu \| \pi M - M \|$ **if consider communities)**

Theorem 2. *Set* $\hat{\pi} = \arg\min ||A - \pi B \pi^T||_F^2$. *If* $||A - \hat{\pi} B \hat{\pi}^T||_F^2 = O(nlogn)$, *then as* n *goes to infinity,* $\frac{||\pi - \pi_0||_F^2}{||\pi_0||_F^2}$ *goes to 0 almost surely.*

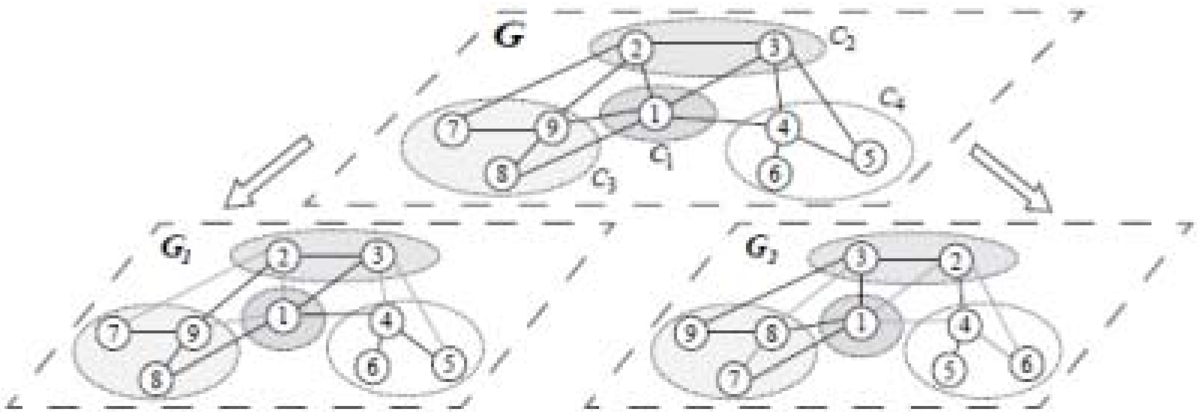
$$||\hat{\pi} - \pi_0||_F^2 \leq \frac{8(1+\eta)}{nps_2(1-s_2)} ||\hat{\pi} B \hat{\pi}^T - A||_F^2 + \frac{16 s_1 n \delta}{1 - s_2}$$

$$\delta = \Theta(logn/n).$$

Node matching error is bounded by Edge matching error, and vanishes as the size goes to infinity.

In fact we find that [min||A- π B π ^{T}||] is bounded by the ordered difference of node degree in two graphs, so if two graphs are similar in node distributions, then they are prone to have smaller edge matching error.

Example of ordered difference of nodes



G1: {3,3,3,2,2,2,1,1,1}

G2: {3,3,2,2,2,2,1,1,0} (|G1|>|G2|)

Ordered difference: {0,0,1,0,0,0,0,0,1}

Contribution 2:

- **Prove that under what condition the edge matching problem has the unique solution.**

Incentives:

- **If there is not unique solution, it is hard to determine which one is best, and it does not benefit the algorithm to find the best solution.**

$$\min \quad ||A\pi - \pi B||_F^2 + \mu||\pi M - M||_F^2$$
$$s.t. \quad \pi \mathbf{1} = \mathbf{1}, \quad \pi^T \mathbf{1} = \mathbf{1}.$$

Theorem 3. *Suppose that graph A and B are isomorphic, and the eigenvalue decomposition of A is $A = U\Lambda U^T$, then if $(\mathbf{1}^T U e_i) U^T M M^T U e_j \neq (U^T \mathbf{1} e_i^T) U^T M M^T U e_j$, then there is a unique solution to the edge matching optimization problem.*

Contribution 3:

- **Modify the problem to accomodate more general situations.**

Incentives:

- **Previous work only considers same number of nodes & communities, but in reality communities and the number of nodes in two graphs are not necessarily the same.**

$$||A\pi - \pi B||_F^2 + \mu||\pi M - M||_F^2$$

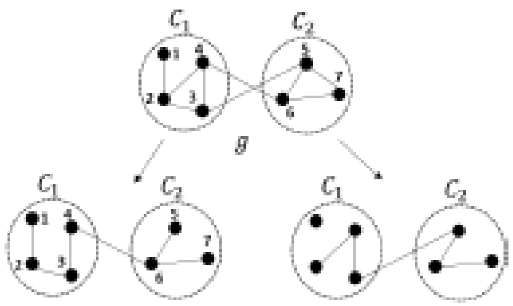
$$F_0(\pi) = ||A - \pi B \pi^T||_F^2 + \mu||\tilde{M} - \pi \tilde{M} \pi^T||_F^2$$

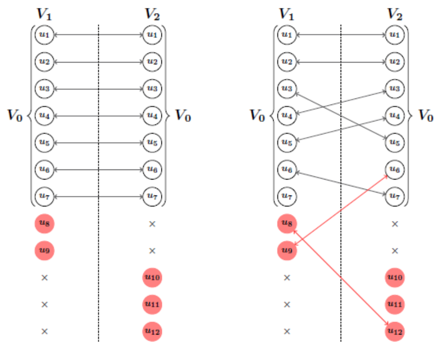
Advantages:

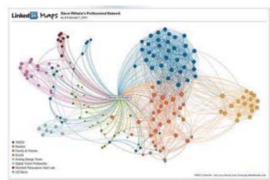
- **More general situations**
- **More symmetric form**

Theorem 4. *Assume that π is an $n \times n$ permutation matrix and there are m communities ($M \in \mathbb{R}^{n \times m}$). For a single row of M, there are $2^m - 1$ different community assignments except the situation that a node does not belong to any community. For a given graph* $\mathbf{G}$, *extract those community assignments* $\mathbf{G}$ *has, and denote them as $C_1, C_2, ..., C_s$ ($0 < s < 2^m - 1$). $|C_i|$ denotes the number of communities the i-th assignment is related to. If there does not exist two assignments C_x and C_y ($x \neq y$), $|C_x| = |C_y|$ and the number of nodes with assignment C_x and C_y is equal, then minimizing $||\pi \tilde{M} \pi^T - \tilde{M}||_F^2$ and minimizing $||\pi M - M||_F^2$ are equivalent*

mild condition







Combined in one form

Contribution 4:

- **Solve the problem by convex-concave technique, derive the algorithm and analyze its convergence.**

Incentives:

- **The edge matching problem is NP-hard, so approximation algorithm is needed. A convex-concave technique avoids the projection process in solely convex relaxation in previous work, which may cause big error.**

Projection:

$$\begin{bmatrix} 1/3 & 1/5 & 1/6 \\ 1/6 & 1/10 & 1/4 \\ 1/10 & 1/7 & 1/8 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Convex-Concave Technique

Convex relaxation: F1(π)

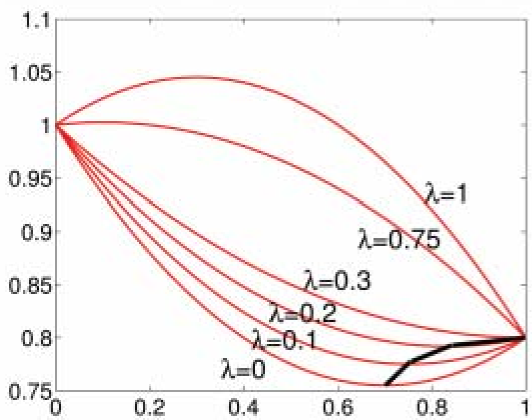
Tractable to find the minimal solution.

Concave relaxation: F2(π)

The minimal solution is on the boundary.

Combination of them:

$$F(\pi) = \lambda F_1(\pi) + (1-\lambda) F_2(\pi)$$



Algorithm 1: Edge and Community Matching Algorithm

Input:

Two adjacent matrices, A and B.
Community assignment matrix, M in B
Weight Controlling parameter μ.

Output:

Estimated permutation matrix $\tilde{\pi}$.

1: Calculate $\tilde{M} = MM^T$
2: Form the objective function $F_0(\pi)$ and its convex-concave relaxation function $F(\pi)$
3: $\xi = 0, \pi = \mathbf{1}_{n \times n}./n$;
4: Calculate the upper limit of ξ as $\xi_m = ||I \otimes A - B \otimes I||_F + \mu||I \otimes \tilde{M} - \tilde{M} \otimes I||_F$
5: **while** $\xi < \xi_m$ and $\pi \notin \Omega_0$ **do**
6:     **while** π not converged **do**
7:         $\pi_{tmp} = \arg\min_\pi \mathbf{tr}(\nabla_\pi F(\pi, \xi)^T \pi)$, where $\pi_{tmp} \in \Omega$
8:         $\gamma = \arg\min_\gamma F(\pi + \gamma(\pi_{tmp} - \pi), \xi)$
9:         Set $\pi = \pi + \gamma(\pi_{tmp} - \pi)$.
10:     **end while**
11:     $\xi = \xi + d\xi$.
12: **end while**

Time Complexity: Calculating ξ_m costs $O(n^4)$. The inner loop is similar to the Frank-Wolfe algorithm. One loop has the complexity $O(n^3)$. Set the maximum number of inner loops is T, thus the outer loop has the complexity $O(n^4 + \frac{n^3 T \xi}{d\xi})$.

Some results:

The convergence rate is strictly 1/k, where k is the iteration time (refer to [3], but modify it into matrix domain)

3. Conclusion & Future Work

Main Contributions:

- **Node matching to Edge matching**
- **Uniqueness**
- **Generalization**
- **Solution and its convergence**

3. Conclusion & Future Work

Future Work:

- The Performance Guarantee of Convex-Concave Algorithm
- Whether the objective function can be derived from MAP or minimum average error criterion?
- Experiment on Real Data

REFERENCES:

[1] X.Fu, et.al. "De-anonymization of Social Networks with Communities:When Quantifications Meet Algorithms".
[2] M.Zaslavskiy, et.al. "A Path Following Algorithm for the Graph Matching Problem".
[3] M.D.Canon and C.D.Cullum, "A Tight Upper Bound on the Rate of Convergence of the Frank-Wolfe Algorithm".

Thanks!

Q&A