Most Influential points in the Social Networks?

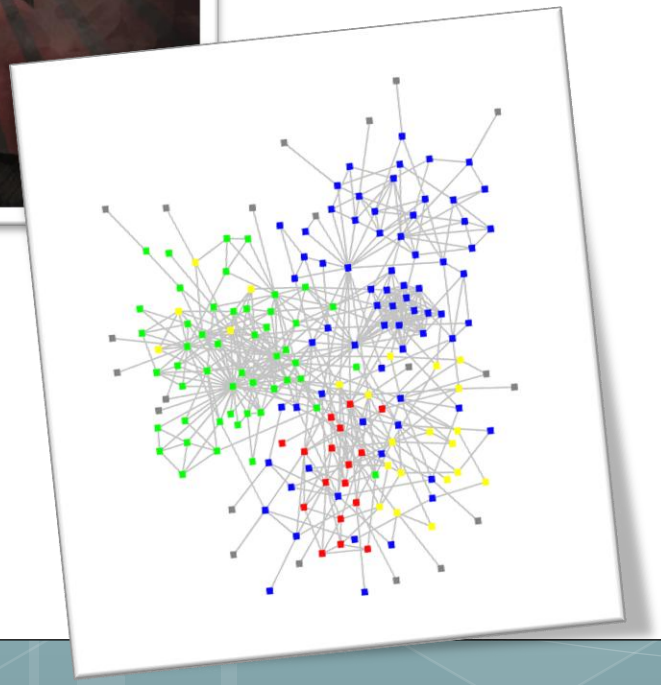
COMP621U Presentation

WANG Guan 王冠 (Crown)
Physics MPhil
April.12th 2011

wangguan@ust.hk

What are the most influential points
In the social networks, or, any networks?

Suppose you are Mega mind,
where would you attack?



wangguan@ust.hk

Outline

- 1. Community Bridges
- 2. Bridge Detection Algorithms
- 3. Proposed New Algorithms:
  Overlapping Community Detection
- Summary

wangguan@ust.hk

1. Community Bridges

[1] M. Salathé and J. H. Jones,

Dynamics and Control of Diseases in Networks with Community Structure

PLoS Comput.Biol 6 (2010)

wangguan@ust.hk

[1] M. Salathé and J. H. Jones, **Dynamics and Control of Diseases in Networks with Community Structure**
PLoS Comput.Biol 6 (2010)

- Question:

How can we best control epidemics (prevention or mitigation)?

How do we get maximum effect with given supply of vaccines?

If we have a targeted immunization strategy, how do we find the targets?

wangguan@ust.hk

The Simplest Epidemic Model: SIR

S = fraction of a population that is susceptible to infection

I = fraction of population that is infectious (will give infection to S on contact)

R = removed or recovered

$$(1) \quad dS/dT = -\beta SI$$

$$(2) \quad dI/dt = \beta SI - \gamma I = I(\beta S - \gamma)$$

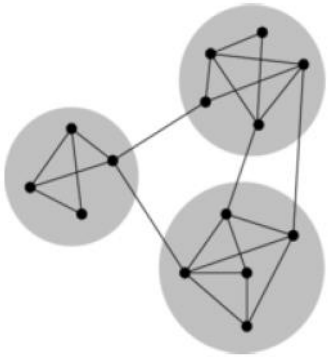
$$(3) \quad dR/dt = \gamma I$$

So, if $S(0) < \gamma/\beta$, then $dI/dt < 0$ and infection dies out.

wangguan@ust.hk

- One can build networks with the same number of nodes, same number of edges, same degree distribution, and **still get fundamentally different epidemic dynamics**.

- Reason: **Community Structures**



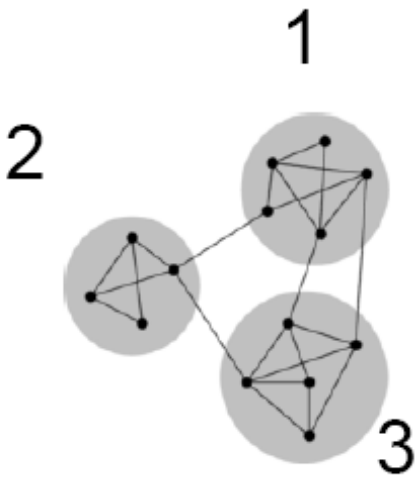
The occurrence of groups of nodes that are more densely connected within a group than between a group. Very common in many networks, especially social networks.

wangguan@ust.hk

Modularity: $$Q = \sum_i \left(e_{ii} - a_i^2 \right), \quad a_i = \sum_j e_{ij}$$

e_{ij} is fraction of all edges in the network that link nodes in community i to nodes in community j.

a_i is fraction of edges in the network that connect to nodes in community i.

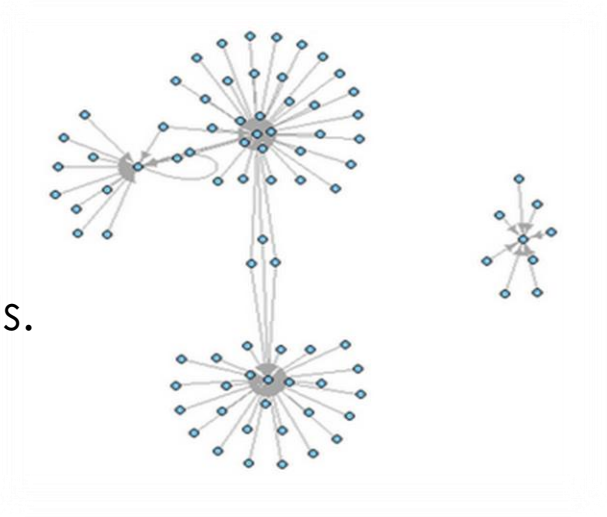


	1	2	3	a	a^2	$e_{ii}-a^2$
1	7/23	0.5/23	1 /23	**8.5/23**	**0.136**	**0.168**
2	0.5 /23	5 /23	0.5 /23	**6/23**	**0.068**	**0.149**
3	1/23	0.5 /23	7 /23	**8.5/23**	**0.136**	**0.168**
				1		**0.486**

wangguan@ust.hk

- Targeting highly connected individuals for immunization is an effective strategy to mitigate epidemics.

Prediction of epidemics based on only degree fails.



- In networks with strong community structure, this may not be effective because some individuals connect to many communities but have low degree:
- These are **community bridges**.

wangguan@ust.hk

- Community Bridge Evaluation:
- **Betweenness Centrality**

  The fraction of shortest paths that a node falls on.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is total number of shortest paths between s, t and $\sigma_{st}(v)$ is the number of those paths that go through v.

How to find such community bridges??

wangguan@ust.hk

2. Bridge Detection Algorithms

- [1] M. Salathé and J. H. Jones, PLoS Comput.Biol 6 (2010).
- [2] M. E. J. Newman, Social networks 27, 39 (2005).
- [3] N. Madar, T. Kalisky, R. Cohen, D. Ben-avraham, and S. Havlin, The European Physical Journal B-Condensed Matter and Complex Systems 38, 269 (2004).

Algorithms knowing Complete Network Structure: [2]
Algorithms without knowing Complete Network Structure: [1][3]

wangguan@ust.hk

- **Algorithms knowing Complete Network Structure**

- Newman's **Random Walk Centrality** algorithms.
- Identify target nodes by a random walk, counting how often a node is crossed by a random walk between two other nodes.
- This includes not only shortest paths but all paths between a pair of nodes, and it gives shorter paths a higher weight.
- **This algorithm is actually used for community detection.**

[2] M. E. J. Newman, Social networks 27, 39 (2005).

wangguan@ust.hk

- **Algorithms without knowing Complete Network Structure**

  - Cohen's **Acquaintance Method** algorithms
  - Fix a number N and identify target nodes by picking a random neighbor of a random node. Once a node (acquaintance) has been picked N times, we say this node is one of the most influential. When N = 1, it means every node is the one of the most influential.
  - Work well in fat tailed (scale-free) networks.

[3] N. Madar, T. Kalisky, R. Cohen, D. Ben-avraham, and S. Havlin
The European Physical Journal B-Condensed Matter and Complex Systems 38, 269 (2004)

wangguan@ust.hk

- **Algorithms without knowing Complete Network Structure cont.**

- Salathe's **Community-bridge-finder (CBF)** algorithms
- Start from a random node and follow a random path on the complex network until it arrives at a node that does not connect back to more than one of the previously visited nodes on the random walk
- Such a node is more likely to be on the "bridge", and such nodes connect to multiple communities.
- The first node that does not connect back to previously visited nodes in the current random walk is likely to be in a different community.
- Here they are talking about social network, and thus each node should have at least more than two neighbors.

[1] M. Salathé and J. H. Jones, PLoS Comput.Biol 6 (2010)

wangguan@ust.hk

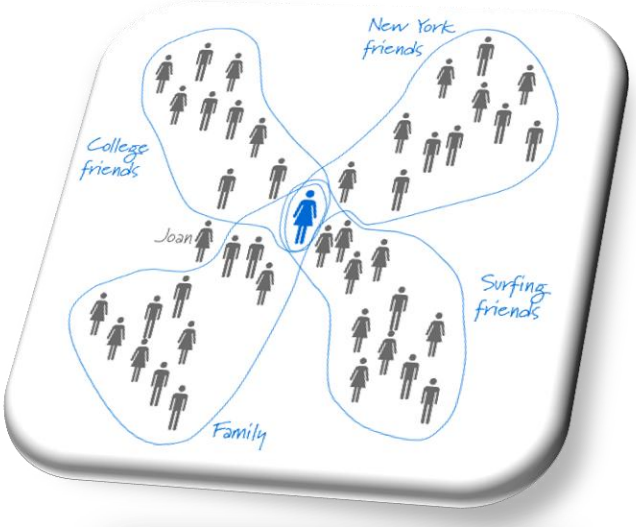
3. Proposed New Algorithms:
Overlapping Community Detection

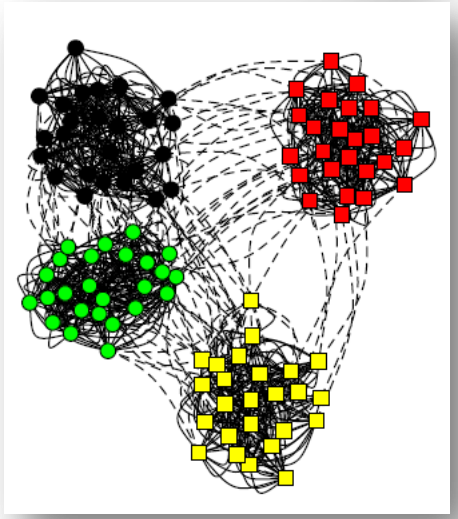
- [4] Xufei Wang, Lei Tang, Huiji Gao, and Huan Liu. **Discovering Overlapping Groups in Social Media**. In Proceedings of The 10th IEEE International Conference on Data Mining (ICDM'10), 2010
- [5] Zhenggang Wang and K. Y. Szeto, **Structure profile of complex networks by a model of precipitation**, Physica A: Statistical Mechanics and its Applications, Volume 389, Issue 11, 1 June 2010, Pages 2318-2324

wangguan@ust.hk

- Motivated by [1], we can have the following statement:

- **The most influential nodes (links) in social networks can be the ones that belong to the largest number of multiple communities, known as "Community Bridges".**





wangguan@ust.hk

Most influential links:

Normal Community Detection!

Once we find the communities of nodes, it's easy to identify the links between those communities.

Links between the biggest communities will be most influential.

wangguan@ust.hk

Most influential nodes:

Community Detection on the links!

Nodes of links which belong to the most number of link communities will be the most influential ones.

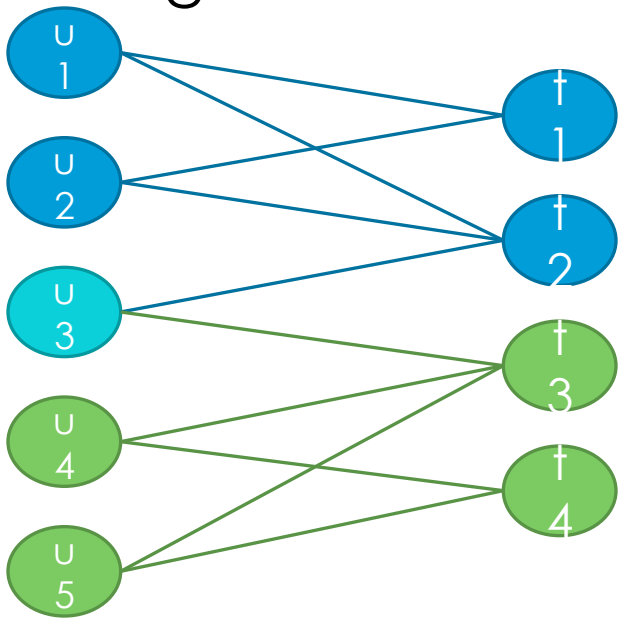
How do we find the **communities of links**?

wangguan@ust.hk

Edge-centric View

- Cluster edges instead of nodes into disjoint groups
  - One node can belong to multiple groups. One edge belongs to one group.

In [4], they focus on a User-Tag subscription, which is a Bipartite Graph.
We can generalize it to **any networks**.

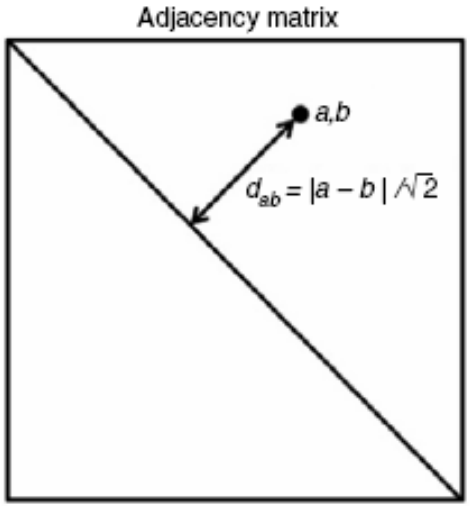


wangguan@ust.hk

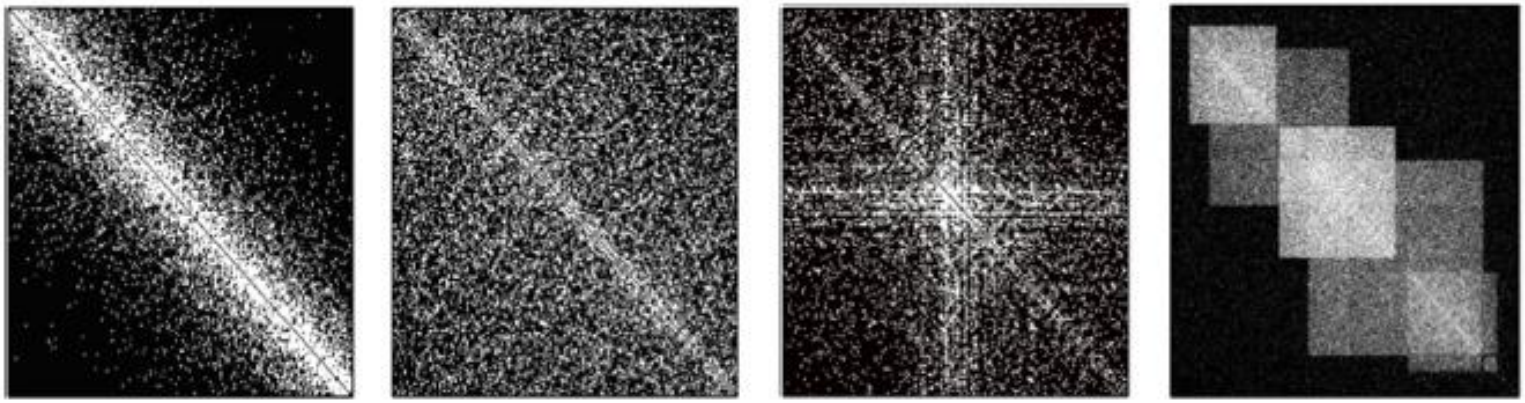
[5] gives an interesting algorithm:

$$H(E) = \sum_{(i,j)\in E} w_{ij} d_{ij} = \frac{1}{\sqrt{2}} \sum_{(i,j)\in E} w_{ij} |i - j|,$$

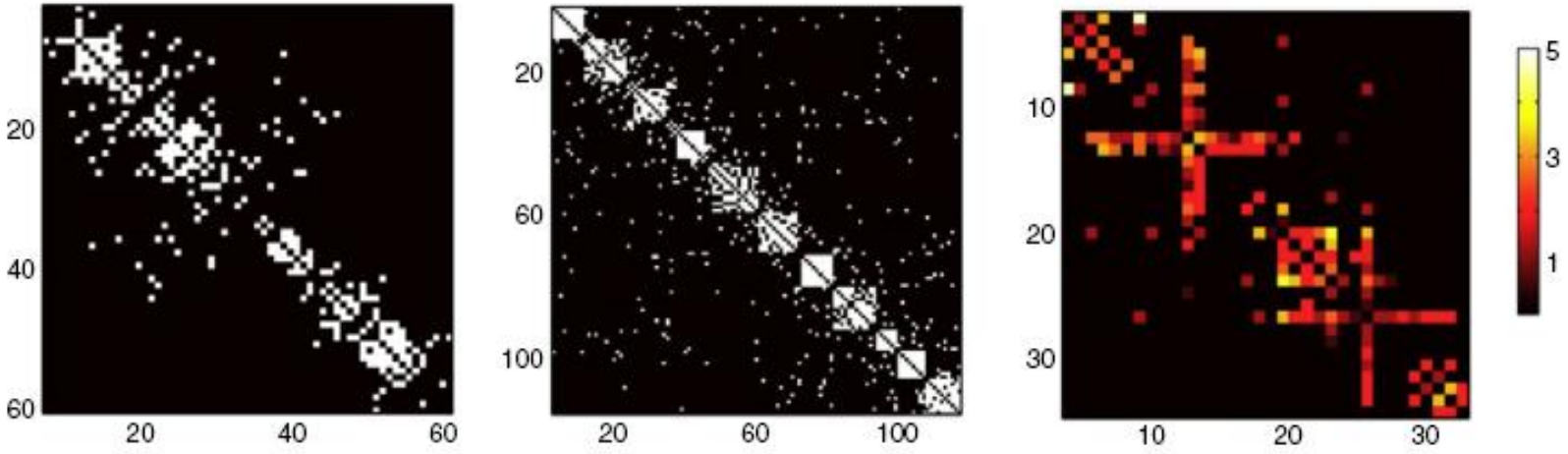
Z. Wang, K.Y. Szeto / Physica A 389 (2010) 2318–2324



wangguan@ust.hk



(a) Artificial. (b) ER network. (c) BA network. (d) Overlapping.



(a) dolphin (b) football (c) karate

- [5] is basically a edge-clustering method. It is able to show nodes that belongs to a pair of communities.

- Question: Can we modify the algorithm to show better results for multiple (>2) overlapping communities?

wangguan@ust.hk

Summary

- 1. Community Bridges
- 2. Bridge Detection Algorithms
- 3. Proposed New Algorithms:
  Overlapping Community Detection

Questions?

Thank you!

wangguan@ust.hk