UMassAmherst

Reciprocity in Social Networks with Capacity Constraints

Bo Jiang[1], Zhi-Li Zhang[2], Don Towsley[1]

[1]UMass Amherst
[2]Univ of Minnesota

MURI, March 6, 2015

Introduction

- Many complex networks are directed
  - WWW, Wikipedia, ...
  - Twitter, Google+, Flickr, LiveJournal, YouTube, ...
- Reciprocity measures tendency to form reciprocal links
  - nontrivial patterns reveal organizational principles
  - observed in many real networks

Twitter(2007)	Twitter(2009)	Google+(2011)	Spanish Wiki
0.55	0.28	0.32	0.35

How to Assess Nontriviality?

Question

Swedish Wiki has reciprocity 21%. Is this nontrivial?

Traditional Answer

- Compare with expected value in null models
  - random graph w/ same # nodes & edges
  - random graph w/ given degree sequence
- Classify as *reciprocal* or *anti-reciprocal*
  - reciprocal if larger than random
  - anti-reciprocal if smaller than random
- For Swedish Wiki, random $\approx 0 \Rightarrow$ reciprocal

But...

Does 21% reciprocity mean strong tendency to reciprocate?

- compared with 0? maybe...
- compared with 100%? not quite...
- what if maximum is 28%? yes!

Lesson

Extremal values are informative & important!

- Focus on maximum reciprocity
    - real social networks have reciprocity larger than random
- Need to solve reciprocity maximization problem

Any Constraints?

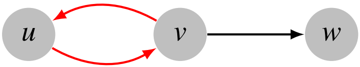
- Degree sequence is key structural feature
  - better be preserved for fair comparison
- Proxy for “capacity” constraints
  - file sharing network (source to downloader)
    - in-degree: bandwidth
    - out-degree: resource
  - social network (follower to followee)
    - in-degree: fame & popularity
    - out-degree: budget of attention
- Preserving degree sequence honors capacity constraints

Reciprocity

- **Defn**: fraction of edges with reciprocal edge

$$r(G) = \frac{\rho(G)}{\varepsilon(G)}$$

  - $\rho(G)$: # reciprocated links
  - $\varepsilon(G)$: total # edges
  - simple digraph, i.e. no self-loops or multiple edges



$$r(G) = \frac{2}{3}$$

Degree Bi-sequence

- Every graph G is associated w/ *bi-sequence* $(\mathbf{d}^+, \mathbf{d}^-)$
  - out-degree seq: $\mathbf{d}^+ = (d_1^+, d_2^+, \ldots, d_n^+)$
  - in-degree seq: $\mathbf{d}^- = (d_1^-, d_2^-, \ldots, d_n^-)$
- *Graphic* bi-sequence: realizable by digraph
  - Not every bi-sequence is graphic
  - Graphicality test: theorems of Erdös-Gallai type
- $\mathcal{G}(\mathbf{d}^+, \mathbf{d}^-)$: set of all digraphs with bi-sequence $(\mathbf{d}^+, \mathbf{d}^-)$
  - $\mathcal{G}(\mathbf{d}^+, \mathbf{d}^-)$ is nonempty $\Leftrightarrow$ $(\mathbf{d}^+, \mathbf{d}^-)$ is graphic

Maximum Reciprocity Problem (MRP)

- Find digraph G in $\mathcal{G}(\mathbf{d}^+, \mathbf{d}^-)$ with maximum $\rho(G)$

$$\begin{array}{ll} \text{maximize} & \rho(G) \\ \text{subject to} & G \in \mathcal{G}(\mathbf{d}^+, \mathbf{d}^-). \end{array}$$

  - for fixed $(\mathbf{d}^+, \mathbf{d}^-)$, max $\rho(G) \Leftrightarrow$ max $r(G)$

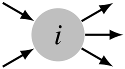
- Call any maximizing G a *maximum (reciprocity) digraph*
- Denote $\rho(\mathbf{d}^+, \mathbf{d}^-) = \max \rho(G)$

Upper Bound

$$\rho(\mathbf{d}^+, \mathbf{d}^-) \leq \sum_i d_i^+ \wedge d_i^- = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$$

- # reciprocated edges leaving i bounded by

$$\rho_i \leq d_i^+ \wedge d_i^-$$



- Necessary condition for equality
  - both $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ are graphic
- Graphicality of $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ can be violated independently

Example 1

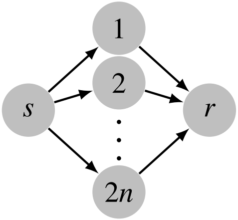
- Neither $\mathbf{d}^+ \wedge \mathbf{d}^-$ nor $\mathbf{d}^+ \vee \mathbf{d}^-$ is graphic, since they have odd sums.

i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
1	$(1,0)$	0	1
2	$(1,1)$	1	1
3	$(0,2)$	0	2
4	$(2,1)$	1	2
5	$(1,1)$	1	1

- $\rho(\mathbf{d}^+, \mathbf{d}^-) = 2 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = 3$

Example 2

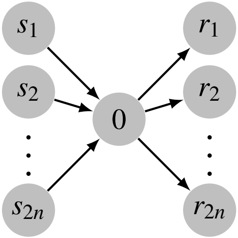
- $\mathbf{d}^+ \vee \mathbf{d}^-$ is not graphic while $\mathbf{d}^+ \wedge \mathbf{d}^-$ is.
- $\rho(\mathbf{d}^+, \mathbf{d}^-) = 0 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = 2n$



i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
s	$(2n, 0)$	0	$2n$
$1 \sim 2n$	$(1, 1)$	1	1
r	$(0, 2n)$	0	$2n$

Example 3

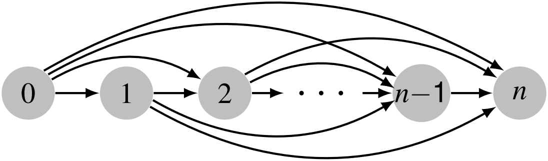
- $\mathbf{d}^+ \wedge \mathbf{d}^-$ is not graphic while $\mathbf{d}^+ \vee \mathbf{d}^-$ is graphic
- $\rho(\mathbf{d}^+, \mathbf{d}^-) = 0 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = 2n$



i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
$s_1 \sim s_{2n}$	$(1, 0)$	0	1
$r_1 \sim r_{2n}$	$(0, 1)$	0	1
0	$(2n, 2n)$	$2n$	$2n$

Insufficiency of necessary condition

- both $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ are graphic when $n \equiv 0 \mod 4$
- $\rho(\mathbf{d}^+, \mathbf{d}^-) = 0 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$



i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
$0 \sim \lfloor n/2 \rfloor$	$(n-i, i)$	i	$n-i$
$\lceil n/2 \rceil \sim n$	$(n-i, i)$	$n-i$	i

MRP is NP-hard

Theorem

It is NP-complete to decide whether $\rho(\mathbf{d}^+, \mathbf{d}^-) = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$.

Proof.

By reduction from 3-color tomography problem. □

Sufficient Condition

Theorem

Assume $\mathbf{d}^+ \vee \mathbf{d}^- > \mathbf{0}$, *i.e. no isolated nodes.*
$\rho(\mathbf{d}^+, \mathbf{d}^-) = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$ *if*

- $\mathbf{d}^+ \wedge \mathbf{d}^-$ *and* $(\mathbf{d}^+, \mathbf{d}^-) - \mathbf{d}^+ \wedge \mathbf{d}^-$ *are graphic;*
- $\Delta < \sqrt{n}$, *where* $\Delta = ||(\mathbf{d}^+, \mathbf{d}^-)||_\infty$

Proof.

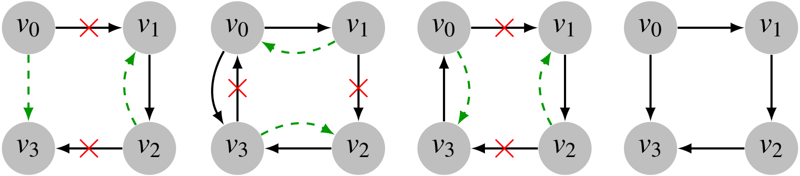
Related to packing of graphic degree sequence. □

Bad News

$\Delta < \sqrt{n}$ usually fails for real networks

Suboptimal Motifs

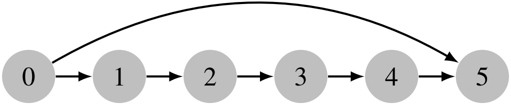
- 3-path: elementary length-3 path of unreciprocated edges
- 4 types, classified by connection between v_0 and v_3
- Types I, II, III are suboptimal



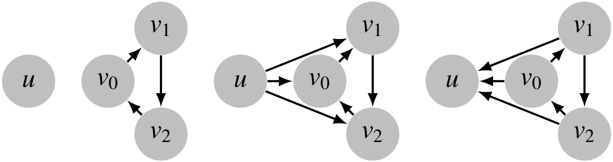
(a) Type I (b) Type II (c) Type III (d) Type IV

3-path Optimal Digraphs

- no 3-path of Type I, II, or III
- can be obtained by repeated rewiring
- properties of subgraph induced by unreciprocated edges
  - odd length elementary paths must have shortcuts

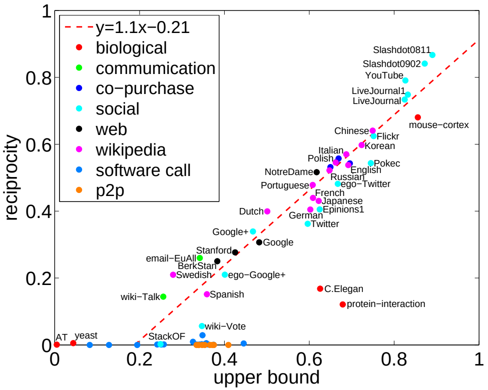


  - no cycles other than vertex-disjoint 3-cycles
  - vertices of 3-cycles have same connectivity to outside



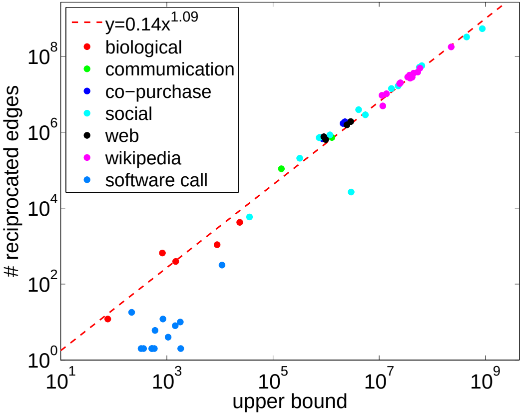
Empirical Study

- reciprocity varies widely
  - Gnutella: 0
  - Slashdot: 90%
  - high for social & Wiki
  - low for software call
- Strong linear relationship



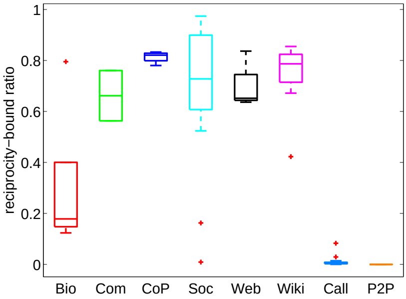
Empirical Study

- # reciprocated edges vs. upper bound



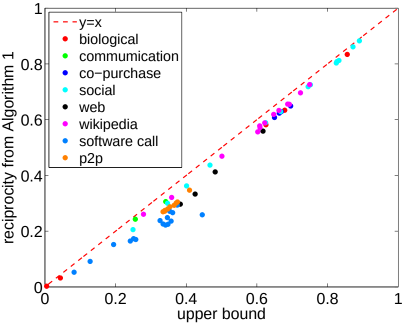
Empirical Study

- reciprocity-bound ratio has much narrower range
- ratio $> 50\%$ for communication, co-purchasing, social and web networks except for
  - wiki-vote
  - Spanish Wikipedia
  - Stack Overflow Q&A
- Strong tendency to reciprocate modulo degree constraints
  - Swedish Wiki: 21% vs. 75%
  - Google+: 34% vs. 73%



Empirical Study

- 3-path optimal digraphs have reciprocities close to upper bounds
- upper bound summaries fundamental limit imposed by degree bi-sequence
- suboptimal 3-paths are major cause of loss in reciprocity



Future Work

- How to estimate/approximate maximum reciprocity?
  - approx. algorithm w/ performance guarantee
- How to estimate reciprocity for large networks?
  - performance of sampling methods
  - streaming algorithm
- Does it help to know degree distributions?
  - heavy-tailed vs. light-tailed distributions
- Other characteristics?
  - clustering coefficient