Impact of Heavy-tailed Joint Degree Distribution on Network Characteristics

Bo Jiang
Joint work by UMass & UMN

MURI, Oct. 7, 2014

Introduction

- Many different graph characteristics of interest
  - degree distributions: in-degree, out-degree, joint, ...
  - reciprocity: fraction of links with reciprocal link
  - clustering coefficient

Network	in-deg	out-deg	recip.	cl. coeff.
Spanish Wiki	PL(1.26)[1]	PL(1.70)[1]	0.3517[2]	0.09[1]
Portuguese Wiki	PL(1.10)[1]	PL(1.80)[1]	0.3563[2]	0.05[1]
Twitter (2007)	PL(1.4)[3]	PL(1.4)[3]	0.58[3]	0.106[3]
Twitter (2009)	PL(1.3)[4]	PL(1.3)[4]	0.22[4]	
BlogPulse (2006)	PL(1.38)[5]	PL(1.5?)[5]	0.033[5]	0.063[5]
Google+ (2011)	PL(1.3)[6]	PL(1.2)[6]	0.32[6]	

[1] Zlatić et. al. 2006
[2] Zamora-López et. al. 2008
[3] Java et. al. 2007
[4] Kwak et. al. 2009
[5] Shi et. al. 2006
[6] Magno et. al. 2012

Introduction

- Different characteristics are correlated
  - specification of one imposes constraints on others
  - less freedom for design, but more power for inference

Introduction

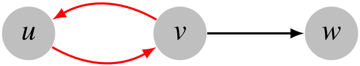
- Different characteristics are correlated
  - specification of one imposes constraints on others
  - less freedom for design, but more power for inference
- Focus on impact of joint degree distribution on reciprocity
  - to what extent is reciprocity determined by joint degree distr.?
  - does reciprocity behave differently for heavy- and light-tailed joint degree distr.?
  - what's impact of in- and out-degree correlation?

Reciprocity

- **Defn**: fraction of links with reciprocal link

$$r = \frac{R}{L}$$

  - # reciprocated links R
  - total # links L
  - simple digraph, i.e. no self-loops



$$r = \frac{2}{3}$$

Reciprocity

- Given graphical bi-degree sequence $(d_1^+, d_1^-), \ldots, (d_N^+, d_N^-)$, what are possible values of r?

Reciprocity

- Given graphical bi-degree sequence $(d_1^+, d_1^-), \ldots, (d_N^+, d_N^-)$, what are possible values of r?
- Configuration model: randomly pair up stubs

Reciprocity

- Given graphical bi-degree sequence $(d_1^+, d_1^-), \ldots, (d_N^+, d_N^-)$, what are possible values of r?
- Configuration model: randomly pair up stubs

- Expected reciprocity is [Zamora-López et .al. 2008]

$$\langle r \rangle = \bar{a} \frac{\langle d^+ d^- \rangle^2}{\langle d \rangle^4}$$

where $\bar{a} = L/N^2$ is link density, $\langle d \rangle = \langle d^+ \rangle = \langle d^- \rangle$.

Correlation & Reciprocity

- Reciprocity can be rewritten

$$\langle r \rangle = \bar{a} \left(\rho c_v^+ c_v^- + 1 \right)^2$$

  - ρ is corr. coeff. between in- and out-degrees.
  - c_v^+, c_v^- are coeff. of variation. for out- and in-degrees

Correlation & Reciprocity

- Reciprocity can be rewritten

$$\langle r \rangle = \bar{a} \left(\rho c_v^+ c_v^- + 1 \right)^2$$

  - ρ is corr. coeff. between in- and out-degrees.
  - c_v^+, c_v^- are coeff. of variation. for out- and in-degrees

- positive corr. $\Rightarrow$ high reciprocity
  negative corr. $\Rightarrow$ low reciprocity
  uncorrelated $\Rightarrow \langle r \rangle = \bar{a}$

Crude Calculation

- Assume $\{d_i^+\}$, $\{d_i^-\}$ have power-law distributions F^+, F^- with respective exponents α^+, α^-
- Use expected order statistics of i.i.d. samples from F^+ to approximate corresponding order statistics of $\{d_i^+\}$
- Similarly for $\{d_i^-\}$
- WLOG, assume $d_1^+ \geq d_2^+ \geq \cdots \geq d_N^+$

Crude Calculation

- Positively correlated: $d_1^- \geq d_2^- \geq \cdots \geq d_N^-$

$$\langle r_{pos} \rangle \approx \bar{a} \frac{(1-\beta^+)^2(1-\beta^-)^2}{(\beta^+ + \beta^- - 1)^2} N^{2(\beta^+ + \beta^- - 1) \vee 0}$$

  where $\beta^+ = 1/\alpha^+$, $\beta^- = 1/\alpha^-$.
- Negatively correlated: $d_1^- \leq d_2^- \leq \cdots \leq d_N^-$,

$$\langle r_{neg} \rangle \approx \bar{a}(1-\beta^+)^2(1-\beta^-)^2 B^2(1-\beta^+, 1-\beta^-)$$

  where $B(\cdot, \cdot)$ is beta function.

Crude Calculation

- Positively correlated: $d_1^- \geq d_2^- \geq \cdots \geq d_N^-$

$$\langle r_{pos} \rangle \approx \bar{a} \frac{(1-\beta^+)^2(1-\beta^-)^2}{(\beta^+ + \beta^- - 1)^2} N^{2(\beta^+ + \beta^- - 1) \vee 0}$$

  where $\beta^+ = 1/\alpha^+$, $\beta^- = 1/\alpha^-$.
- Negatively correlated: $d_1^- \leq d_2^- \leq \cdots \leq d_N^-$,

$$\langle r_{neg} \rangle \approx \bar{a}(1-\beta^+)^2(1-\beta^-)^2 B^2(1-\beta^+, 1-\beta^-)$$

  where $B(\cdot,\cdot)$ is beta function.
- e.g. for $\alpha^+ = \alpha^- = 1.5$, $N = 10^6$,

$$\langle r_{pos} \rangle \approx 1111\ \bar{a}$$
$$\langle r_{neg} \rangle \approx 0.3468\ \bar{a}$$

  Note $\bar{a} \to 0$ in sparse network.

What if Less Random?

- Average is not enough, reality may be less random
- Comparison with extremals may be informative of regularity

What if Less Random?

- Average is not enough, reality may be less random
- Comparison with extremals may be informative of regularity
- Simple bounds:

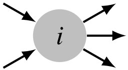
$$0 \leq r \leq \frac{\sum_i d_i^+ \wedge d_i^-}{\sum_i d_i^+}$$

  - total # links

$$L = \sum_i d_i^+$$

  - # reciprocated links leaving i bounded by

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



Crude Estimation of Upper Bound

- Positively correlated: $d_1^- \geq d_2^- \geq \cdots \geq d_N^-$

$$UB_{pos} \approx 1 + \left(\frac{1-\beta^+}{1-\beta^-}\right)^{\frac{1-\beta^-}{\beta^+-\beta^-}} - \left(\frac{1-\beta^+}{1-\beta^-}\right)^{\frac{1-\beta^+}{\beta^+-\beta^-}}$$

- Negatively correlated: $d_1^- \leq d_2^- \leq \cdots \leq d_N^-$

$$UB_{neg} \approx 2 - \left[(1-\gamma)^{1-\beta^-} + \gamma^{1-\beta^+}\right]$$

where γ satisfies $(1-\beta^+)(1-\gamma)^{1-\beta^-} = (1-\beta^-)\gamma^{\beta^+}$.

Numeric Example

For $\alpha^+ = \alpha^- = 1.5$, $N = 10^6$,

- average

$$\langle r_{pos} \rangle \approx 1111\bar{a}$$
$$\langle r_{neg} \rangle \approx 0.3468\bar{a}$$

- upper bound

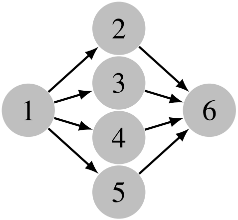
$$UB_{pos} \approx 1$$
$$UB_{neg} \approx 0.413$$

Bound is Loose

- Bi-degree sequence

i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$
1	$(4, 0)$	0
$2 \sim 5$	$(1, 1)$	1
6	$(0, 4)$	0

- Upper bound $r \leq 1/2$
- Actual r must be *zero*



Questions

- For given bi-degree sequence, what's maximum achievable reciprocity?
- Is upper bound asymptotically tight for power-law graphs?

Upper Bound Revisited

- # reciprocated links

$$R \leq \sum_i d_i^+ \wedge d_i^- = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$$

where $\mathbf{d}^+ = (d_1^+, \ldots, d_N^+)$, $\mathbf{d}^- = (d_1^-, \ldots, d_N^-)$.

Upper Bound Revisited

- # reciprocated links

$$R \leq \sum_i d_i^+ \wedge d_i^- = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$$

  where $\mathbf{d}^+ = (d_1^+, \ldots, d_N^+)$, $\mathbf{d}^- = (d_1^-, \ldots, d_N^-)$.
- Necessary condition for equality
  - both $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ are graphical

Upper Bound Revisited

- # reciprocated links

$$R \leq \sum_i d_i^+ \wedge d_i^- = ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1$$

  where $\mathbf{d}^+ = (d_1^+, \ldots, d_N^+)$, $\mathbf{d}^- = (d_1^-, \ldots, d_N^-)$.
- Necessary condition for equality
  - both $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ are graphical
- Graphicality of $\mathbf{d}^+ \wedge \mathbf{d}^-$ and $\mathbf{d}^+ \vee \mathbf{d}^-$ can be violated independently

Some Examples

Example 1

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

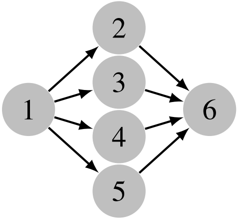
i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
1	$(1,0)$	0	1
2	$(1,1)$	1	1
3	$(0,2)$	0	2
4	$(2,1)$	1	2
5	$(1,1)$	1	1

- $R_{\max} = 2 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = 3$

Some Examples

Example 2

- $\mathbf{d}^+ \vee \mathbf{d}^-$ is not graphical while $\mathbf{d}^+ \wedge \mathbf{d}^-$ is.
- $R_{\max} = 0 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = 4$

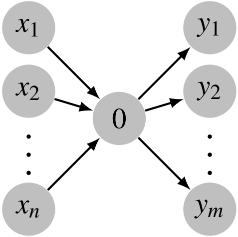


i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
1	$(4,0)$	0	4
$2 \sim 5$	$(1,1)$	1	1
6	$(0,4)$	0	4

Some Examples

Example 3

- $\mathbf{d}^+ \wedge \mathbf{d}^-$ is not graphical while $\mathbf{d}^+ \vee \mathbf{d}^-$ is graphical for even $m \wedge n$.
- $R_{\max} = 0 < ||\mathbf{d}^+ \wedge \mathbf{d}^-||_1 = m \wedge n$



i	(d_i^+, d_i^-)	$d_i^+ \wedge d_i^-$	$d_i^+ \vee d_i^-$
$x_1 \sim x_n$	$(1, 0)$	0	1
$y_1 \sim y_m$	$(0, 1)$	0	1
0	(m, n)	$m \wedge n$	$m \vee n$

Graph with Maximum Reciprocity

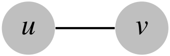
What's structure of graph achieving maximum reciprocity for given bi-degree sequence?

- necessary and sufficient conditions?
- algorithm to maximize reciprocity?

Proposition 1

Suppose $\mathbf{d}^+ = \mathbf{d}^-$*.*

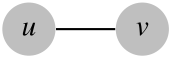
- *Total # links is even* $\Rightarrow R_{\max} = L$*.*



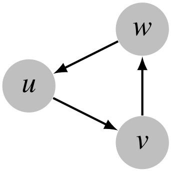
Proposition 1

Suppose $\mathbf{d}^+ = \mathbf{d}^-$.

- *Total # links is even* $\Rightarrow R_{\max} = L$.



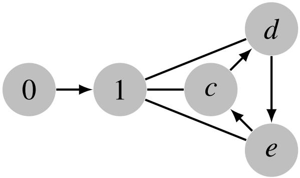
- *Total # links is odd* $\Rightarrow R_{\max} = L - 3$.



Proposition 2

Suppose $d_0^+ - d_0^- = 1$ and $d_1^- - d_1^+ = 1$, $d_i^+ = d_i^-$ for $i \geq 2$.

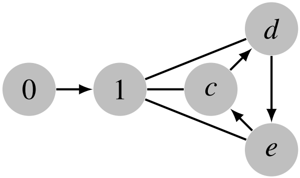
- *Total # links is even $\Rightarrow R_{\max} = L - 2$ or $R_{\max} = L - 4$.*



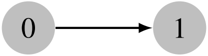
Proposition 2

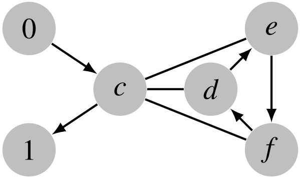
Suppose $d_0^+ - d_0^- = 1$ *and* $d_1^- - d_1^+ = 1$, $d_i^+ = d_i^-$ *for* $i \geq 2$.

- *Total # links is even* $\Rightarrow R_{\max} = L - 2$ *or* $R_{\max} = L - 4$.



- *Total # links is odd* $\Rightarrow R_{\max} = L - 1$ *or* $R_{\max} = L - 5$, *and* $R_{\max} = L - 1$ *iff both* $\mathbf{d}^+ \wedge \mathbf{d}^-$ *and* $\mathbf{d}^+ \vee \mathbf{d}^-$ *are graphical.*





Sufficient Condition

Proposition 3

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

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

Sufficient Condition

Proposition 3

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

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

Question

- If marginal in- and out-degree distr. follow power-law with exponents > 2, then $\Delta < \sqrt{N}$ with high prob.
- What if exponents $\in (1, 2]$?

Slight Generalization

Proposition 4

$R_{\max} \geq L - m$ *if there exists a sequence* $\mathbf{d}^0$ *such that*

- $\mathbf{d}^0$ *and* $(\mathbf{d}^+ - \mathbf{d}^0, \mathbf{d}^- - \mathbf{d}^0)$ *are graphical;*
- $||\mathbf{d}^+ - \mathbf{d}^0||_1 \leq m$*;*
- $\Delta < \sqrt{\delta N + \left(\delta - \frac{1}{2}\right)^2} + \frac{3}{2} - \delta$*, where* $N = |V_0|$*,*
  $\Delta = \bigvee_{i \in V_0}(d_i^+ + d_i^- - d_i^0)$ *and* $\delta = \bigwedge_{i \in V_0}(d_i^+ + d_i^- - d_i^0)$*, with*
  $V_0 = \{i : d_i^+ \vee d_i^- > 0\}$*.*

Question

- How to find $\mathbf{d}^0$?

Ongoing & Future Work

- Better characterization of maximal graph
- More general sufficient condition
- Application to graphs with heavy-tailed degree distr.
- Algorithms to maximize reciprocity
- Application to real network data analysis, e.g. Google+