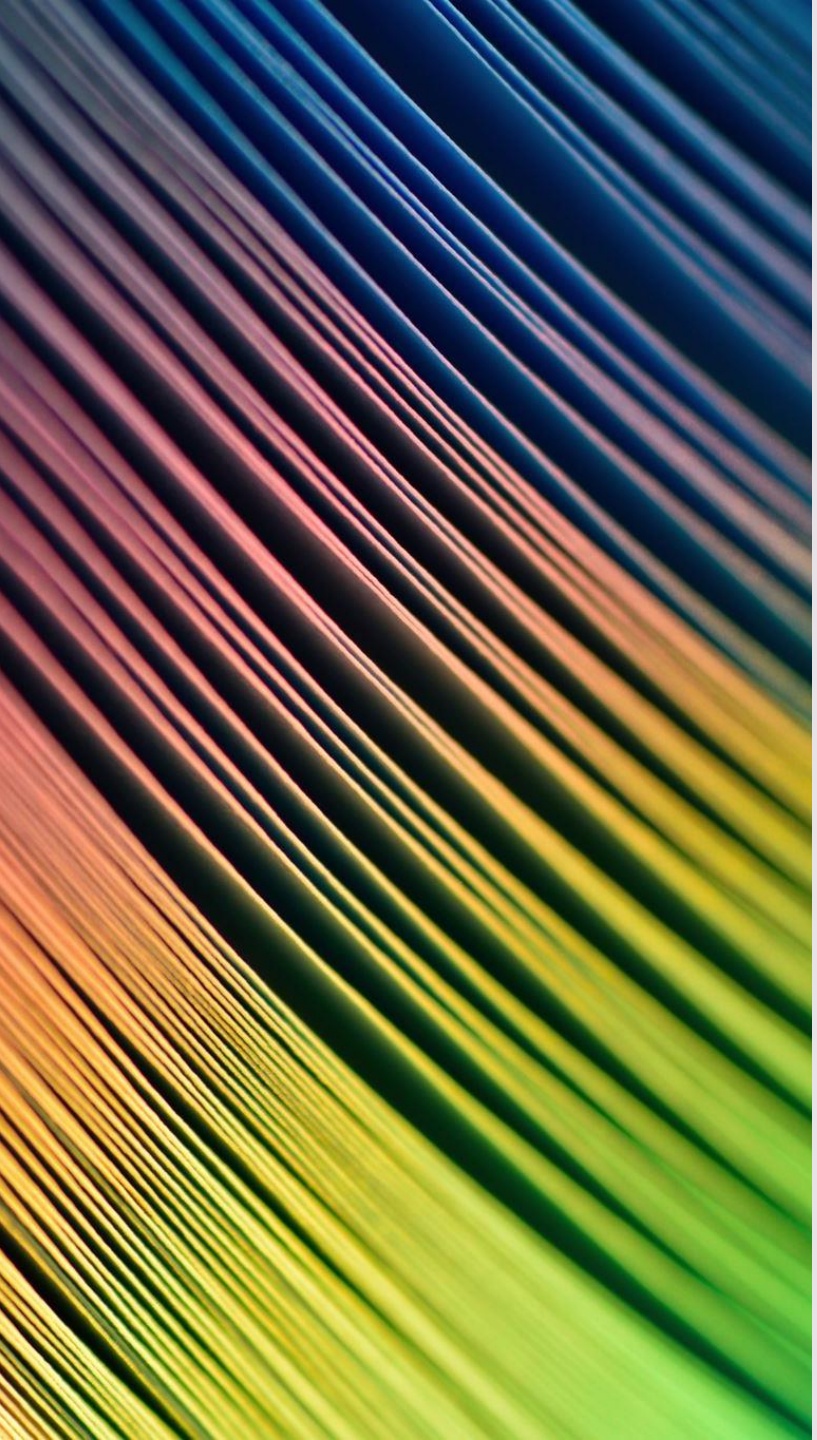


Signed Networks in Social Media

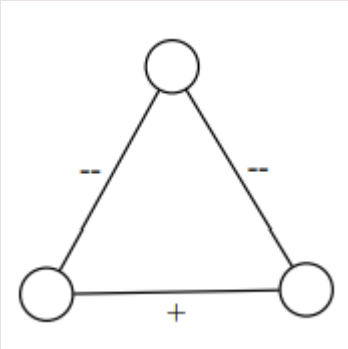
Jure Leskovec

Daniel Huttenlocher

Jon Kleinberg

연구 개요

- Relations between users on social media sites often reflect a mixture of positive (friendly) and negative (antagonistic) interactions.
- Researches has focused almost exclusively on positive interpretations of links between people
- Source : on-line social networks
- ‘Signed Network’



연구 개요

- classical theory : ‘Structural Balance’
- capture certain common patterns of interaction
- at odds with some of the fundamental phenomena we observe (evolving, directed)
- Alternate theory

-> one of the first large-scale evaluations of theories of signed networks using on-line datasets

-> perspective for reasoning about social media sites

연구의 필요성

- Structure of networks in social computing applications
- offers insights into patterns of interactions
- reveals global phenomena at scales that may be hard to identify when looking at a finer-grained resolution
- Richness(complex) -> Stylized network representation(simple) stylized

접근 방법

- theories from social psychology (social balance, social status .etc)

  -> analyzing signed networks of social computing applications

  -> characterize the differences between the observed and predicted configurations

연구에 필요한 요소

- large-scale datasets from social applications where the sign of each link can be reliably determined
- theories of signed networks (Balance, Status)

Datasets

Explicit

- **Epinions (product review) : trust or distrust**
- **Slashdot (blog) : “friends” or “foes**

Implicit

- **Wikipedia admin candidates : public votes in favor of or against**

Edges in networks derived from the datasets are directed

Datasets

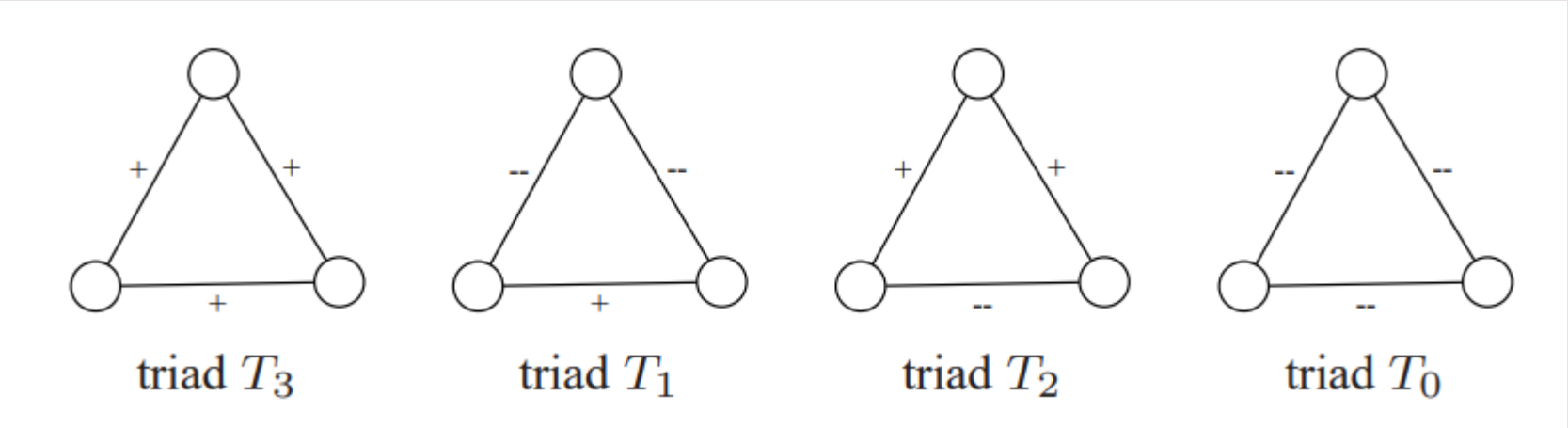
	Epinions	Slashdot	Wikipedia
Nodes	119,217	82,144	7,118
Edges	841,200	549,202	103,747
+ edges	85.0%	77.4%	78.7%
− edges	15.0%	22.6%	21.2%
Triads	13,375,407	1,508,105	790,532

Table 1. Dataset statistics.

Background proportion

Theories of signed networks: Balance

- Initially made for undirected networks
- triangles on three individuals can be signed



- balanced (T3, T1), unbalanced (T2, T0)

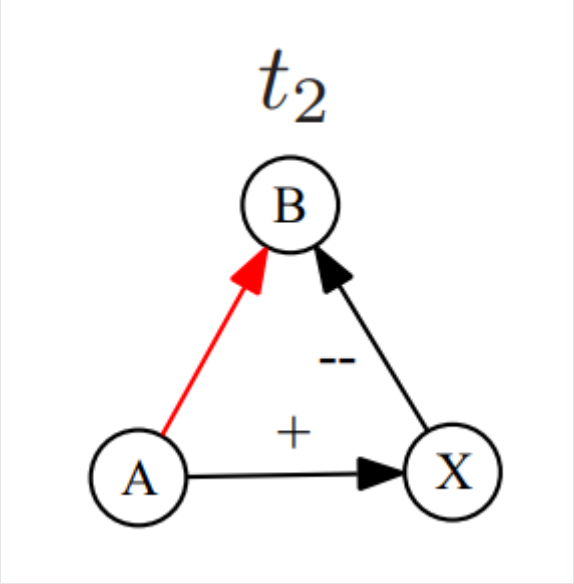
Theories of signed networks: Balance

- **Weak Structural Balance**
- **eliminating - the enemy of my enemy is my friend**
- **only triangles with exactly two positive edges are implausible in real networks**

- **balanced (T3, T1, T0), unbalanced (T2)**

Theories of signed networks: Status

- **Directed networks of signed links**
- **a positive(negative) directed link : the recipient as having higher(lower) status**

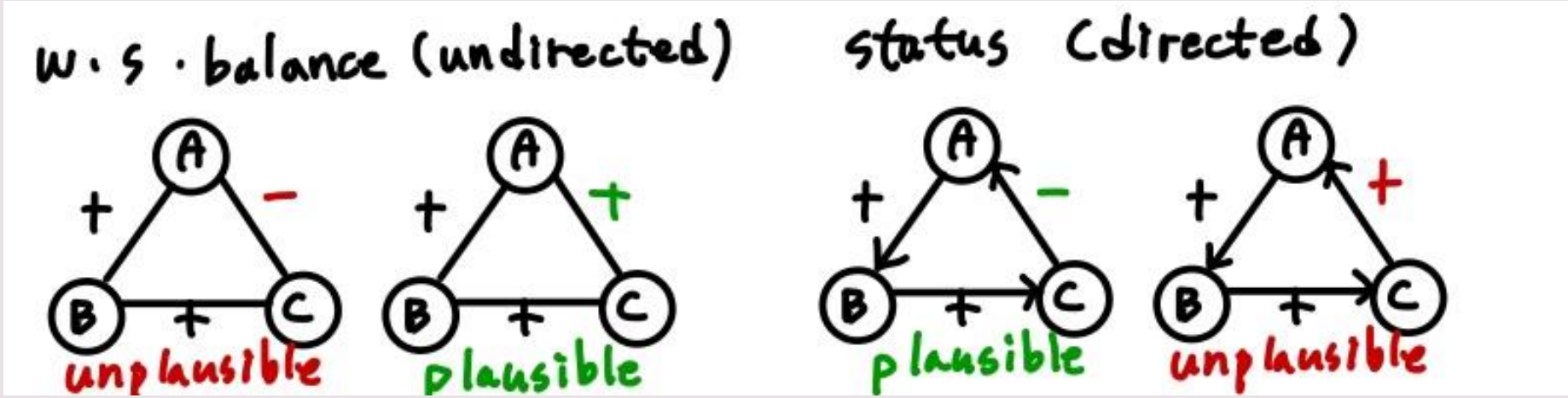


Comparing the two theories

a user A links positively to a user B, and B in turn links positively to a user C

Theories of signed networks: Balance

- (Weak Structural Balance)
- two positive edges are massively underrepresented
- triangles with three positive edges are massively overrepresented
- difficult to evaluate the predictions on large network datasets



추가적인 요소

several experimental conditions

- **including both directed and undirected networks**
- **Respecting and ignoring the order in which edges were created**

Heterogeneity

- **Generative / Receptive**

연구의 결론

- **Comparison of balance and status provides insights into ways in which people use linking mechanisms in social computing applications**

  : to be used more dominantly for expressions of status (Epinions, Wiki)

- **Contrast between balance and status - distinction between undirected and directed interpretations of links**

  : different theories to be appropriate to different levels of resolution in the representation of a single network

연구의 결론 (insight)

- **balance hold more strongly on the subset of links in these networks that are reciprocated (small proportion)**

  : reciprocal link follow different pattern of use

- **connection between the sign of a link and the extent to which it is embedded**

  : many common neighbors (on display) -> more likely to be positive

Symbol	Meaning
T_i	Signed triad, also the number of triads of type T_i
Δ	Total number of triads in the network
p	Fraction of positive edges in the network
$p(T_i)$	Fraction of triads T_i, $p(T_i) = T_i/\Delta$
$p_0(T_i)$	A priori prob. of T_i (based on sign distribution)
$E[T_i]$	Expected number of triads T_i, $E[T_i] = p_0(T_i)\Delta$
$s(T_i)$	Surprise, $s(T_i) = (T_i - E[T_i])/\sqrt{\Delta p_0(T_i)(1-p}$

Table 2. Table of symbols.

ANALYSIS OF UNDIRECTED NETWORKS

- p0 : shuffle signs of all edges
- p(Ti) > p0(Ti) : over-rep

  -> T3
- p(Ti) < p0(Ti) : under-rep

  -> T2

Triad T_i		$\lvert T_i \rvert$	$p(T_i)$	$p_0(T_i)$	$s(T_i)$
Epinions					
T_3	$+++$	11,640,257	0.870	0.621	1881.1
T_1	$+--$	947,855	0.071	0.055	249.4
T_2	$++-$	698,023	0.052	0.321	-2104.8
T_0	$---$	89,272	0.007	0.003	227.5
Slashdot					
T_3	$+++$	1,266,646	0.840	0.464	926.5
T_1	$+--$	109,303	0.072	0.119	-175.2
T_2	$++-$	115,884	0.077	0.406	-823.5
T_0	$---$	16,272	0.011	0.012	-8.7
Wikipedia					
T_3	$+++$	555,300	0.702	0.489	379.6
T_1	$+--$	163,328	0.207	0.106	289.1
T_2	$++-$	63,425	0.080	0.395	-572.6
T_0	$---$	8,479	0.011	0.010	10.8

Table 3. Number of balanced and unbalanced undirected triads.

ANALYSIS OF EVOLVING DIRECTED NETWORKS

1. links are created at specific points in time – order in which links are added

A -> X -> B (semi-path) // A – B (shortcut)

2. different users make use of positive and negative signs differently

(mix, almost exclusively one sign)

-> generative baseline / receptive baseline

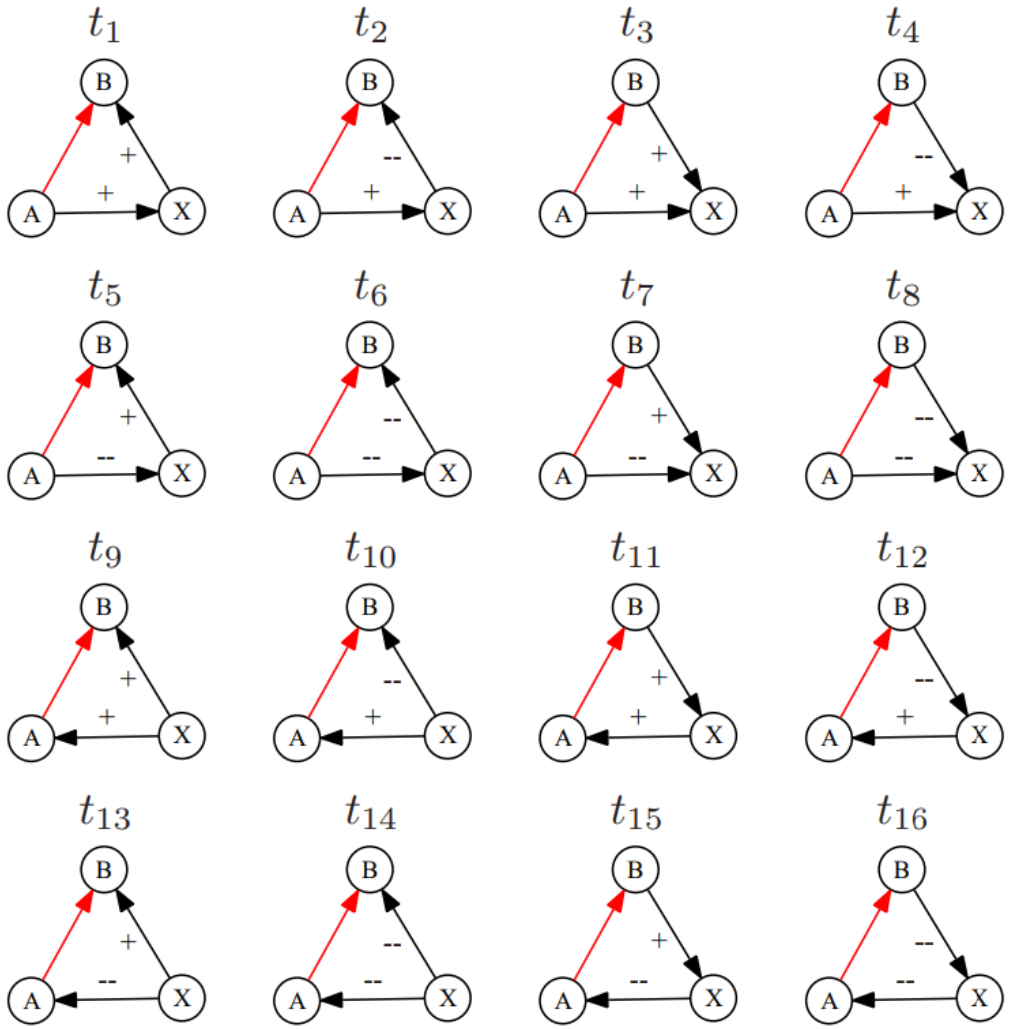
Motivating Example (context)

- 축구 팀에 선수 A와 B 그리고 어떤 동료 X가 있다
- 인터뷰를 통해 X에게 A와 B가 자신(X) 보다 잘하는지 묻는다 (status)
- Directed Graph
- 이때, X가 A도, B도 모두 자신보다 잘한다고 말했다면? (못한다고 했을 때도 동일)
- B는 X보다 잘하므로, 평균보다 높은 status 를 가질 확률이 크다. 따라서, A는 무작위로 한 동료에게 link를 생성할 때보다 B에게 positive 한 link를 생성할 가능성이 높다.
- A의 경우에서 봐도 마찬가지

Contextualized Links (c-links)

t_i	count	$P(+)$	s_g	s_r	B_g	B_r	S_g	S_r
t_1	178,051	0.97	95.9	197.8	✓	✓	✓	✓
t_2	45,797	0.54	-151.3	-229.9	✓	✓	✓	○
t_3	246,371	0.94	89.9	195.9	✓	✓	○	✓
t_4	25,384	0.89	1.8	44.9	○	○	✓	✓
t_5	45,925	0.30	18.1	-333.7	○	✓	✓	✓
t_6	11,215	0.23	-15.5	-193.6	○	○	✓	✓
t_7	36,184	0.14	-53.1	-357.3	✓	✓	✓	✓
t_8	61,519	0.63	124.1	-225.6	✓	○	✓	✓
t_9	338,238	0.82	207.0	-239.5	✓	○	✓	✓
t_{10}	27,089	0.20	-110.7	-449.6	✓	✓	✓	✓
t_{11}	35,093	0.53	-7.4	-260.1	○	○	✓	✓
t_{12}	20,933	0.71	17.2	-113.4	○	✓	✓	✓
t_{13}	14,305	0.79	23.5	24.0	○	○	✓	✓
t_{14}	30,235	0.69	-12.8	-53.6	○	○	✓	○
t_{15}	17,189	0.76	6.4	24.0	○	○	○	✓
t_{16}	4,133	0.77	11.9	-2.6	✓	○	✓	○
			Number of correct predictions		8	7	14	13

P(+): prob. That closing red edge is positive;
sg: surprise of edge initiator giving a positive edge;
sr: surprise of edge destination receiving a positive edge;
Bg: consistency of balance with generative surprise;
Br: consistency of balance with receptive surprise;
Sg: consistency of status with generative surprise;
Sr: consistency of status with receptive surprise.



Result

- the predictions of status with respect to both generative and receptive surprise perform much better than the predictions of structural balance
- status is consistent with generative (14) and receptive (13) surprise on the vast majority of c-link types
- Failed on t3 -> structural balance (f-f-f)
- constitute natural “duals” to each other (reverse the direction and the sign)

  (t3 and t14, t2 and t15)

Result

- Reciprocal (8%) -> structural balance

Epinions	Count	Probability
$P(+\|+)$	38,415	0.969
$P(-\|+)$	1,204	0.031
$P(+\|-)$	1,192	0.692
$P(-\|-)$	560	0.308
Wikipedia	Count	Fraction
$P(+\|+)$	2,509	0.945
$P(-\|+)$	145	0.055
$P(+\|-)$	193	0.706
$P(-\|-)$	80	0.294

Epinions	Triads	P(RSS)	$P(+\|+)$	$P(-\|-)$
Balanced	348,538	0.929	0.941	0.688
Unbalanced	74,860	0.788	0.834	0.676
Wikipedia	Triads	P(RSS)	$P(+\|+)$	$P(-\|-)$
Balanced	53,973	0.912	0.934	0.336
Unbalanced	13,542	0.661	0.878	0.195

B-A link is significantly have the same sign as the A-B link

How public the edge signs are (epinions < Wikipedia)

Directed edges point both -> reinforcing mutual relationship

Embeddedness

- Cluster (closed triad)
- all-positive networks have significantly higher clustering than their randomized counterparts, and the all-negative networks have significantly lower clustering
- ,both the all-positive and all-negative networks are less well-connected than expected

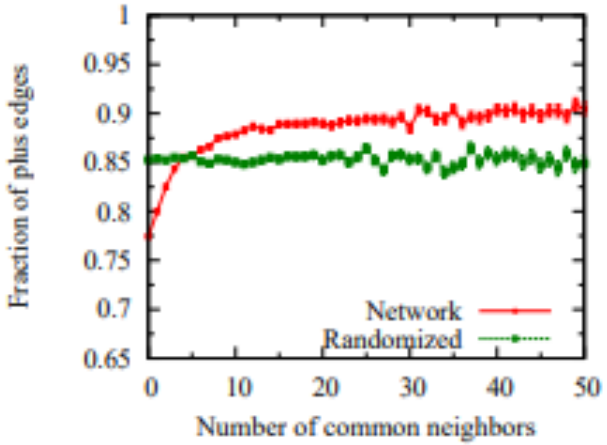
	Size		Clustering		Component	
	Nodes	Edges	Real	Rnd	Real	Rnd
Epinions: −	119,090	123,602	0.012	0.022	0.308	0.334
Epinions: +	119,090	717,027	0.093	0.077	0.815	0.870
Slashdot: −	82,144	124,130	0.005	0.010	0.423	0.524
Slashdot: +	82,144	425,072	0.025	0.022	0.906	0.909
Wikipedia: −	7,115	21,984	0.028	0.031	0.583	0.612
Wikipedia: +	7,115	81,705	0.130	0.103	0.870	0.918

Table 6. Networks composed of only positive (negative) edges. *Real:* network induced on the positive (negative) edges. *Rnd:* network where edge signs are randomly permuted. *Clustering:* fraction of closed triads (closed triads divided by number of length 2 paths) *Component:* fraction of nodes in the largest connected component.

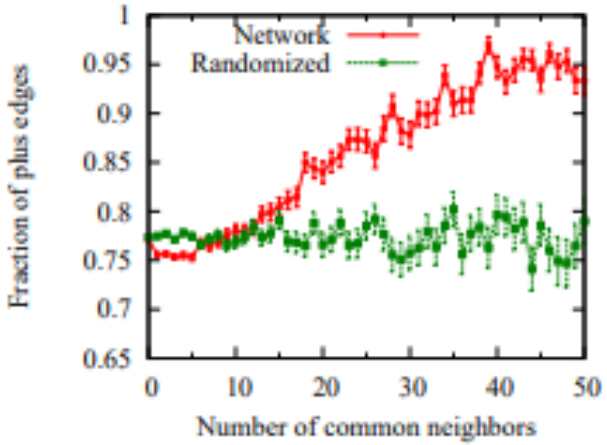
Embeddedness

- **The giant components of real social networks are believed to consistof densely connected clusters linked by less embedded ties**

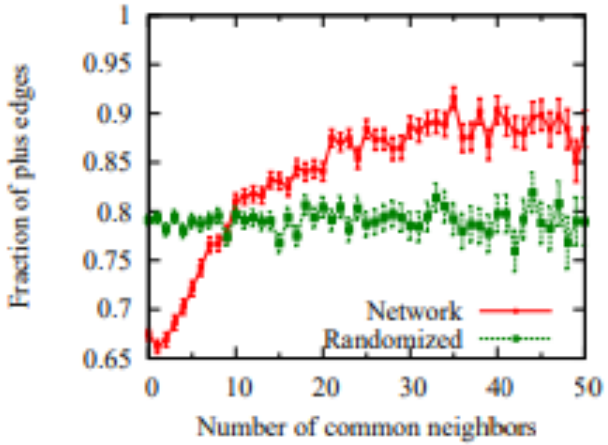
Embeddedness



(a) Epinions



(b) Slashdot



(c) Wikipedia

Figure 3. Embeddedness of positive ties in the network. More embedded edges tend to be more positive.

감사합니다