The Effects of Restrictions on Number of Connections in OSNs
A Case-Study on Twitter

Saptarshi Ghosh　　　　**Gautam Korlam**　　　　**Niloy Ganguly**

Indian Institute of Technology, Kharagpur, India

3rd **Workshop on Online Social Networks**

Boston, USA

June 22, 2010

Restrictions in OSNs

- Most popular OSNs impose restrictions on the number of friends / connections that a member can have

- First line of defence against spam: prevent establishing friendship links with indiscriminately large number of legitimate users

- Reduce strain on the system: most OSNs allow real-time communication from a user to all her friends

- 'Soft' cut-off imposed by Twitter in contrast to fixed or 'hard' limits applied by most OSNs

Motivation of analyzing restrictions

- Restrictions often criticised as encroachment on the freedom of users to have more friends

- Required to design effective restrictions: analysis of the effects of different forms of restrictions on the link-creation dynamics in OSNs

The Restriction in Twitter

The Twitter social network

- Twitter users communicate through the exchange of 'tweets': tweets posted by a user made available to all her followers

- Twitter users form a directed social network: user u 'follows' user v if u is interested in tweets posted by v
  - Nodes: Twitter users
  - Edges: $u \rightarrow v$ if member u follows member v
  - Out-degree of $u \Leftrightarrow u$'s social activity or her interest to collect information from other members
  - In-degree of $u \Leftrightarrow$ popularity of u in the Twitter social network

Follow Spam in Twitter

- Growing popularity of Twitter since 2008 has attracted the attention of spammers

- Many Twitter users engage in 'Aggressive Following' or 'Follow spam'

- "Follow spam is the act of following mass numbers of people, not because you're actually interested in their tweets, but simply to gain attention, get views of your profile (and possibly clicks on URLs therein), or (ideally) to get followed back." [2]

The Twitter Follow-limit

- August 2008: Twitter restricted the number of users that a user can follow (i.e. out-degree) to curb follow-spam and reduce strain on the website [1]

- Every user is allowed to follow up to 2000 others, but "Once you've followed 2000 users, there are limits to the number of additional users you can follow: this limit is different for every user and is based on your ratio of followers to following."

- "Limits improve site performance by ensuring that when we send a person's message to all of their followers, the sending of that message is meaningful."

The Twitter Follow-limit (contd.)

- Twitter does not specify the restriction fully in public

- "We don't reveal exact limits, because it's somewhat complicated and, more importantly, if you were to tell spammers exactly what the filtering rules are on your email or, say, Google's PageRank, they'd just engineer their way around them much more easily." [2]

Conjectures on Twitter Follow-limit

- u_{in}: number of followers (in-degree) of user u

  u_{out}^{max}: maximum number of members whom u can herself follow (maximum possible out-degree)

- version 1: $u_{out}^{max} = \max\{2000,\ 1.1 \cdot u_{in}\}$

- version 2: $u_{out}^{max} = \begin{cases} 2000 + 0.1 \cdot u_{in} & \text{if } u_{in} < 2000 \\ 1.1 \cdot u_{in} & \text{if } u_{in} \geq 2000 \end{cases}$

- Basically, if a user wants to follow (out-degree) more than 2000, she needs to have at least a certain number of followers (in-degree) herself

- Version 1 much more stringent compared to version 2

Experiments on Twitter and Observations

Data Collection using Twitter API

- Challenges
  - Twitter social network has grown too large to collect the entire network
  - Twitter allows at most 150 API calls per hour
- Breadth-first search used to collect 1 million nodes during October 23 - November 8, 2009.
  - Information collected for each user: #friends, #followers, #tweets posted, date of creation of the account, geographical location, ...
- Several smaller crawls starting from randomly selected nodes, during different dates; degree distributions of samples found to be stable irrespective of starting point and time

Scatter plot of followers / friends spread
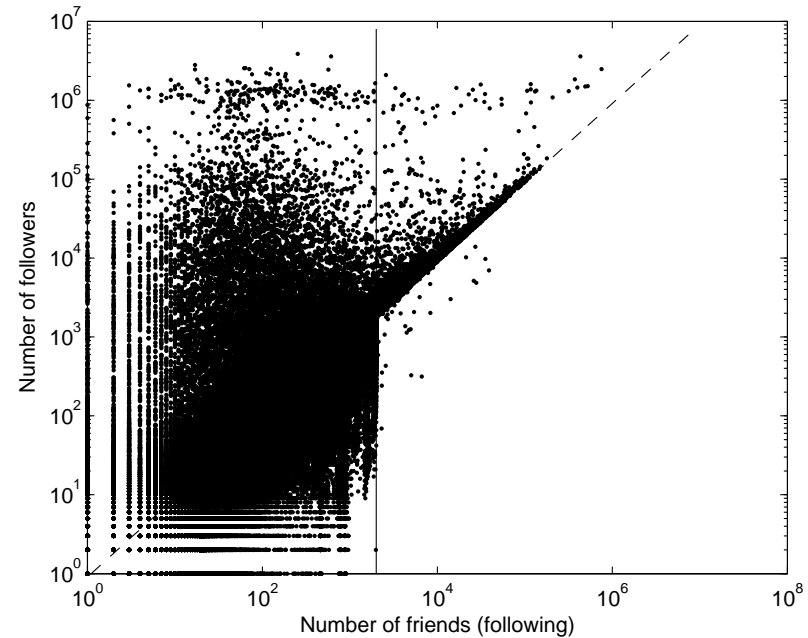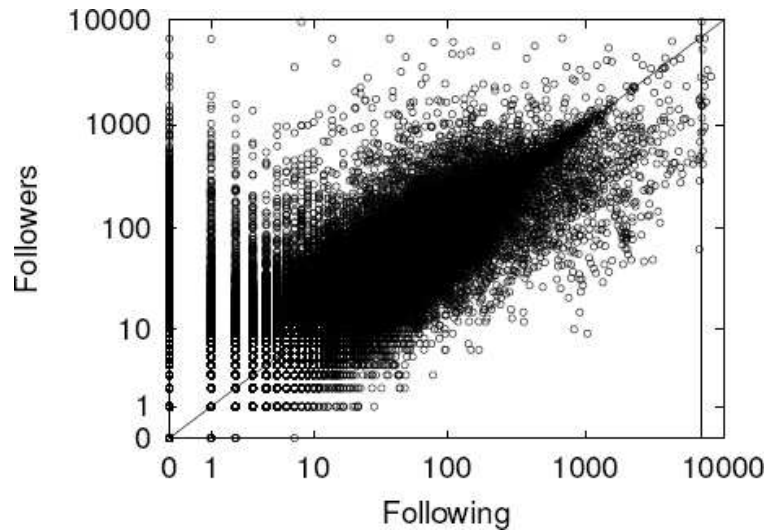
- (left) In Jan-Feb 2008, reproduced from [4]

- (right) In Oct-Nov, 2009 (after restriction imposed)
  - very few users have > 2000 friends (about 6.68%)
  - most users having > 2000 friends lie left of the $x = 1.1 \cdot y$ line: #friends $\leq 1.1 \cdot$ #followers

Degree Distributions
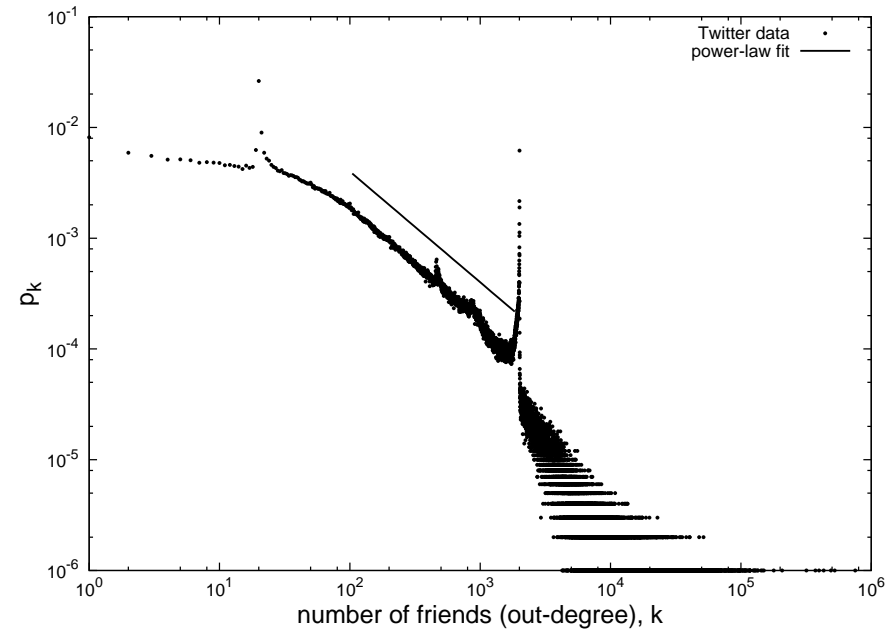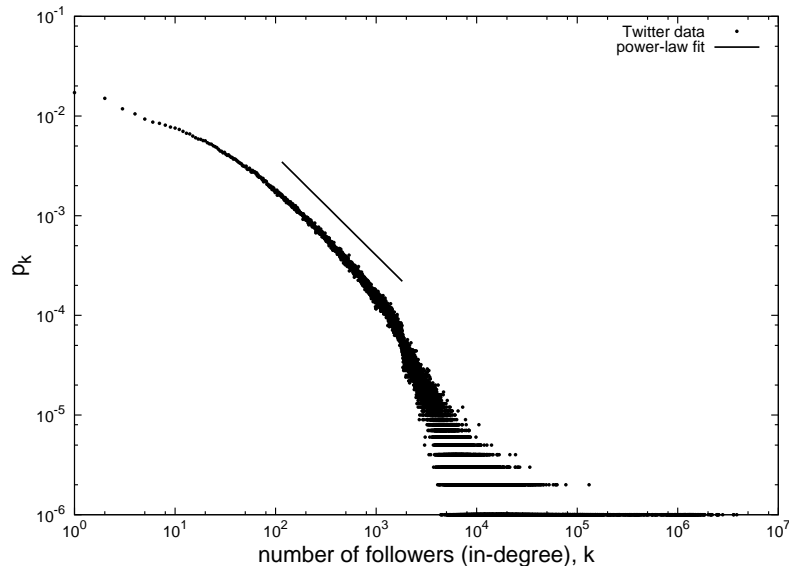
- (left) in-degree distribution   (right) out-degree distribution

- both show power-law fits $p_k \sim k^{-1.0}$ for $k < 2000$

- sharp spike in out-degree distributon around 2000 $\Rightarrow$ a significant fraction of members unable to increase their number of friends beyond a certain limit near 2000

Motivation of analyzing restrictions

- Restrictions often criticised as an encroachment on the freedom of users to have more friends

- Requirements to design effective restrictions: Analysis of the effects of different forms of restrictions on the link-creation dynamics in OSNs

- Topological properties of OSNs can change significantly due to imposed restrictions on node-degree

- Formulate an analytical framework to study the effects of such restrictions on the degree-distribution of a network

Modeling restricted growth dynamics of OSNs

Preferentiality in link dynamics

- Preferential creation of links

  - Members create new links in proportion to their current out-degree

  - A member already having many out-links (friends) is socially more active, hence more likely to create more out-links

- Preferential reception of links

  - Members receive new links in proportion to their current in-degree

  - A member who already has many in-links (followers) is a popular member, hence more likely to get new followers

Model by Krapivsky et. al.

- We customize a growth model [3] for directed networks by incorporating restrictions on degree

- At each time step, one of the following events occurs:

  - Event 1: with probability p, a new node introduced

  - Event 2: with probability $q = 1 - p$, a new directed edge $u \rightarrow v$ created between two existing nodes

Model (contd.)

- Event 1: with probability p, a new node u introduced

  - u forms a directed out-edge to an existing node v

  - Probability of a particular v being selected $\propto (v_{in} + \lambda)$

  - New member u is more likely to follow a popular member v

- Event 2: with probability $q = 1 - p$, a new directed edge $u \rightarrow v$ created between two existing nodes

  - Probability of a particular $u \rightarrow v$ edge $\propto (u_{out} + \mu)(v_{in} + \lambda)$

  - A socially active member u is more likely to follow another member v, especially if v is popular herself

- λ, μ: model parameters that introduce randomness in preferential rules

Model (contd.)

- $N_{ij}(t)$: average number of nodes with in-degree i, out-degree j at time t

- Rate of change in $N_{ij}(t)$ due to change in in-degree of nodes:

$$\frac{dN_{ij}}{dt}_{in} = \left[\frac{(i-1+\lambda)N_{i-1,j} - (i+\lambda)N_{ij}}{I + \lambda N} \right]$$

- Rate of change in $N_{ij}(t)$ due to change in out-degree of nodes:

$$\frac{dN_{ij}}{dt}_{out} = q \left[\frac{(j-1+\mu)N_{i,j-1}\beta_{ij} - (j+\mu)N_{ij}\beta_{i,j+1}}{J + \mu N} \right]$$

- The total rate of change in the number N_{ij} of (i,j)-nodes is

$$\frac{dN_{ij}}{dt} = \frac{dN_{ij}}{dt}\bigg|_{in} + \frac{dN_{ij}}{dt}\bigg|_{out} + p\delta_{i0}\delta_{j1}$$

- Last term accounts for introduction of new nodes with in-degree 0 and out-degree 1

Incorporating restrictions in the model

- Restrictions $\Rightarrow$ only a fraction of the existing nodes can create new out-links

- β_{ij} = 1 iff members having in-degree i are allowed to have out-degree j

- Can be defined to model a variety of restrictions

- Notations used to specify different generalized restrictions:
    - k_c: out-degree at which the restriction starts (2000 in Twitter)
    - 'α-percent rule' (α = 10 in Twitter)

The Twitter Follow-limit (recap)

- u_{in}: number of followers (in-degree) of user u

  u_{out}^{max}: maximum number of members whom u can herself follow (maximum possible out-degree)

- version 1 (known as the '10% rule'):

$$u_{out}^{max} = \mathsf{max}\{2000,\ 1.1 \cdot u_{in}\}$$

- version 2:

$$u_{out}^{max} = \begin{cases} 2000 + 0.1 \cdot u_{in} & \text{if } u_{in} < 2000 \\ 1.1 \cdot u_{in} & \text{if } u_{in} \geq 2000 \end{cases}$$

Modeling different restrictions

- For version 1:

$$\beta_{ij} = \begin{cases} 1 & \text{if } j \leq \max\{ k_c, (1 + \frac{1}{\alpha})i \}, \forall i \\ 0 & \text{otherwise} \end{cases}$$

- For version 2:

$$\beta_{ij} = \begin{cases} 1 & \text{if } i < k_c \text{ and } j \leq k_c + \frac{1}{\alpha}i \\ 1 & \text{if } i \geq k_c \text{ and } j \leq (1 + \frac{1}{\alpha})i \\ 0 & \text{otherwise} \end{cases}$$

- For a 'hard' cut-off at out-degree k_c:

$$\beta_{ij} = \begin{cases} 1 & \text{if } j \leq k_c, \forall i \\ 0 & \text{otherwise} \end{cases}$$

Significance of the model parameters

- p: controls the relative number of nodes and edges (network density)
  - the average in-degree and average out-degree are both $1/p$
  - density of OSNs known to vary over time [5]

- λ, μ: how closely the dynamics of link-formation follow preferential attachment
  - preferential attachment may increase due to recommendation of popular members to new members (as done in Twitter)
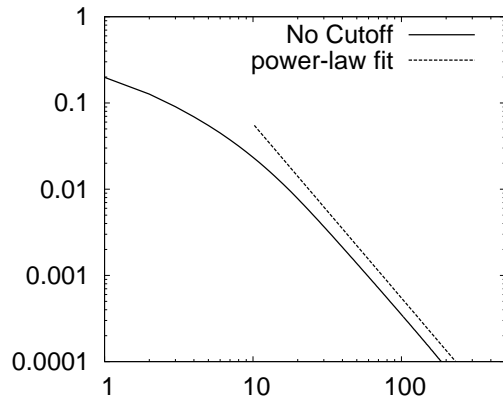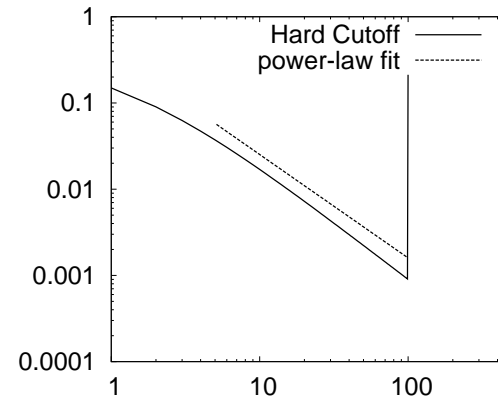
Results from the model

Validating the model

- Proposed theoretical model validated by stochastic simulation

- Parameters: $p = 0.01$, $\lambda = \mu = 1.0$, $k_c = 50$, $\alpha = 10$

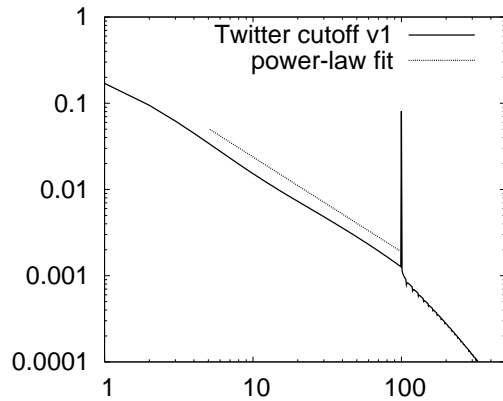- Exact agreement of the simulation results with theory
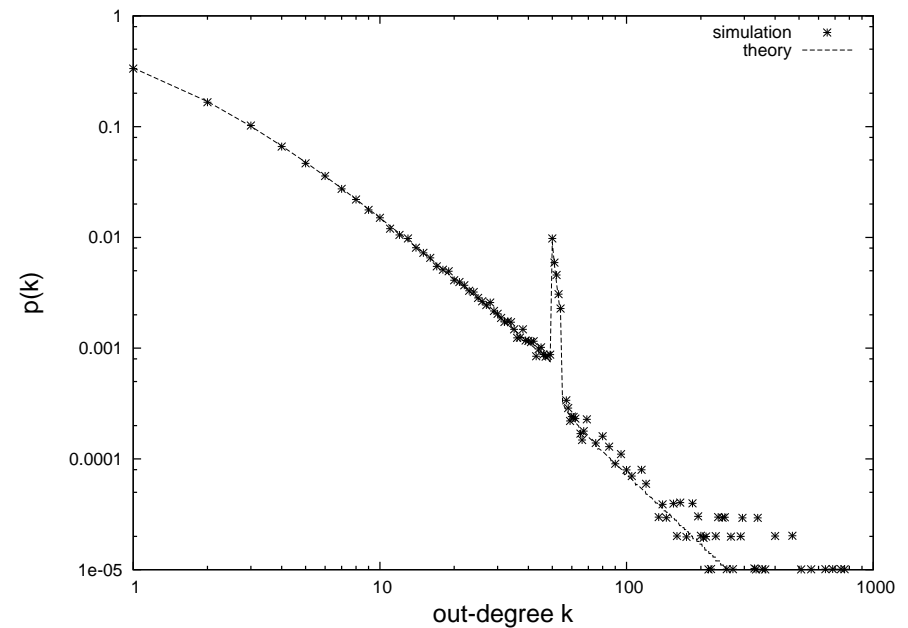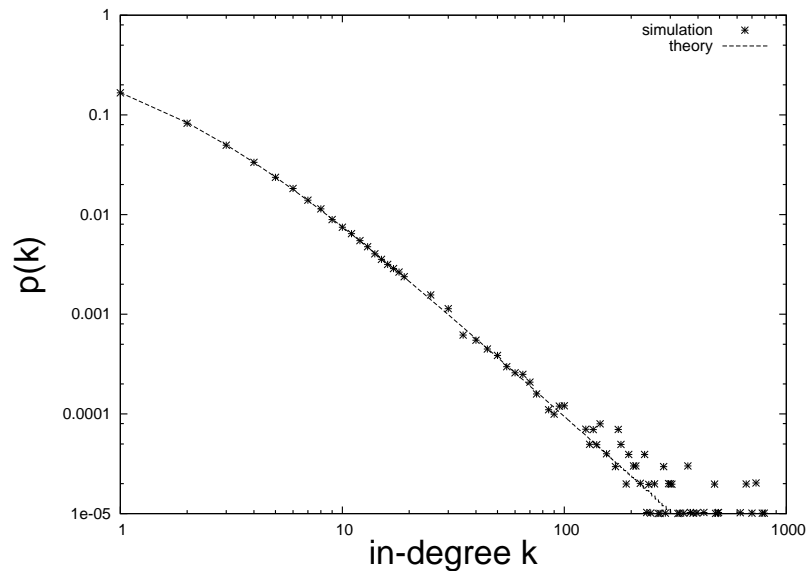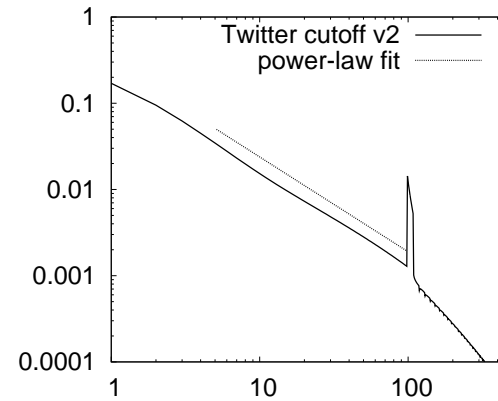
Different Types of Restrictions

No restriction

Hard cutoff

Twitter restriction version 1

Twitter restriction version 2

# Different Types of Restrictions (contd.)

- Both 'hard' and 'soft' restrictions reduce the absolute value of the power-law exponent

- Smaller $|\gamma|$ indicates a more homogeneous structure of the network w.r.t. degrees $\Rightarrow$ reduces strain on hubs

Effects of the network dynamics

- Fraction of nodes that cross the restriction (out of all nodes) measured for different $\lambda = \mu$ and p

  - Increases rapidly with λ (= μ) for their lower values, but stabilizes for higher values of λ (= μ)

  - Reduces sharply with increase in p signifying lesser activity and more growth
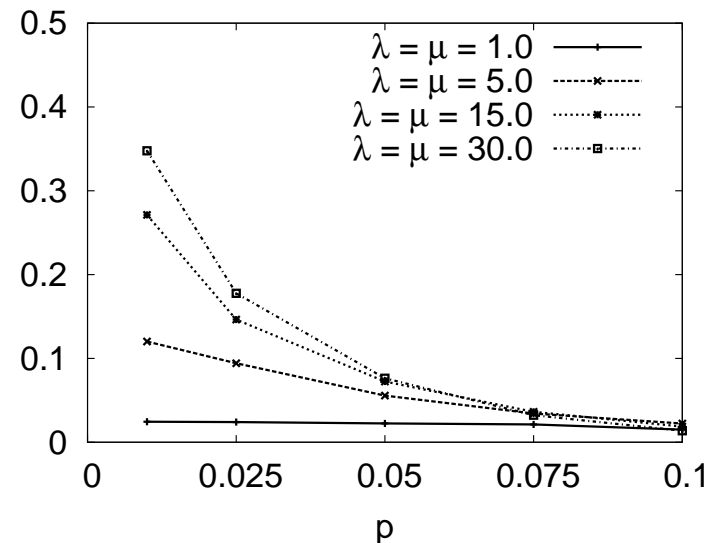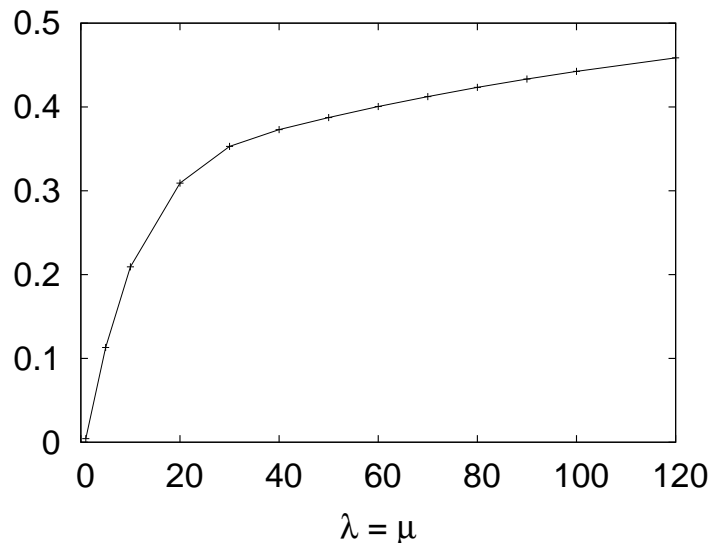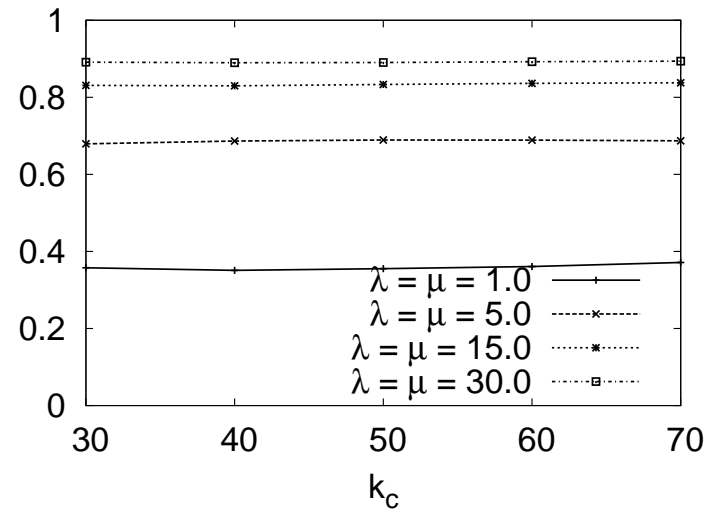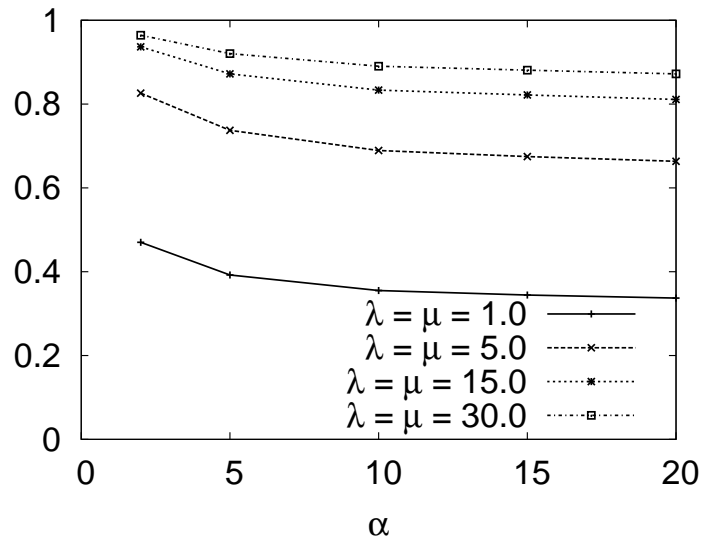
Choice of cut-off parameters
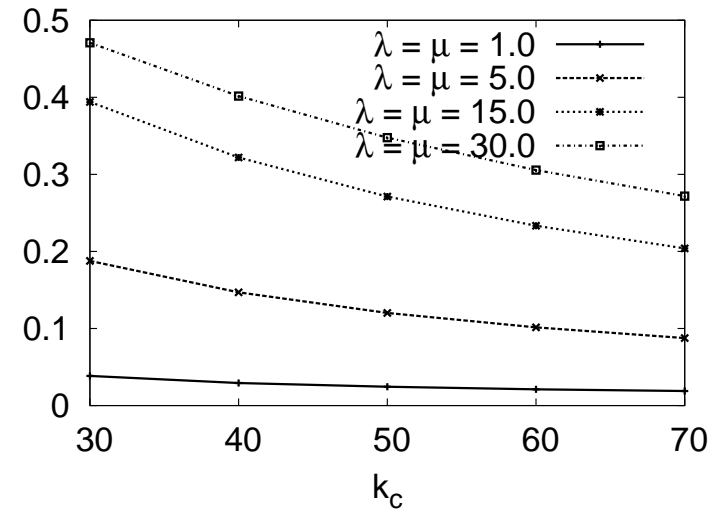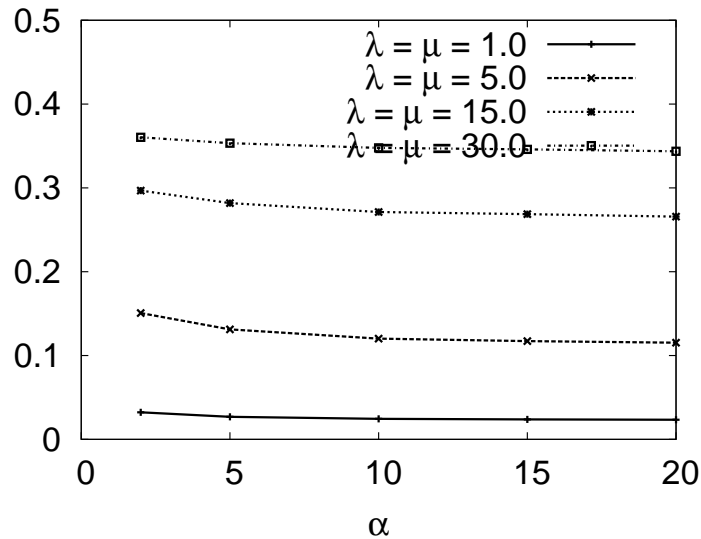
- Number of nodes which cross Twitter cut-off (version 1), measured as a fraction of total number of nodes

- Different values of $\lambda = \mu$ in the range 1.0 to 30.0

  - Does not change appreciably with α

  - Falls rapidly with increase in k_c, for more random dynamics (relatively higher values of $\lambda = \mu$)

Choice of cut-off parameters (contd.)

- Number of nodes which cross Twitter cut-off (version 1), measured as a fraction of the number of nodes which approach the cut-off

- Different values of $\lambda = \mu$ in the range 1.0 to 30.0

  - relatively invariant with k_c

  - reduces with the increase in α, especially in the range $\alpha < 10$

Conclusions drawn from the model

- Preferentiality hinders users from crossing the restriction

- Role of different restriction parameters
  - Importance of k_c: to limit the fraction of members in the whole network, that are able to cross an imposed cut-off
  - Importance of α: more effective in deciding what fraction of the members who approach the cut-off are able to overcome it

- Proposed model can also be used to design restrictions with varying levels of difficulty in overcoming them

References

[1] Twitter support: Following limits and best practices.
`http://help.twitter.com/forums/10711/entries/68916.`

[2] Twitter blog: Making progress on spam.
`http://blog.twitter.com/2008/08/making-progress-on-spam.html,`
August 2008.

[3] KRAPIVSKY, P. L., RODGERS, G. J., AND REDNER, S. Degree distributions of growing networks.
Phys. Rev. Lett. 86, 23 (Jun 2001), 5401–5404.

[4] KRISHNAMURTHY, B., GILL, P., AND ARLITT, M. A few chirps about twitter. In *WOSN '08* (2008),
ACM, pp. 19–24.

[5] KUMAR, R., NOVAK, J., AND TOMKINS, A. Structure and evolution of online social networks. In
KDD (2006), ACM, pp. 611–617.

Thank You

Complex Network Research Group

Department of CSE, IIT Kharagpur, India

http://www.cse-web.iitkgp.ernet.in/~cnerg/