UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Information Diffusion: Influential Neighbours Selection in Online Social Networks

Eiko Yoneki
Systems Research Group
University of Cambridge Computer Laboratory

with Hyoungshick Kim
UBC

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

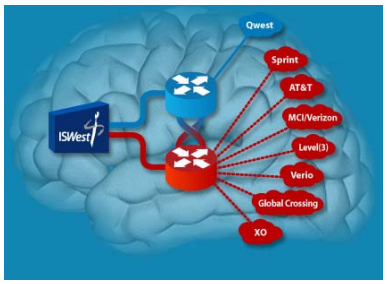
Outline

- What is EU RECOGNITION Project?

  http://www.recognition-project.eu/

- Influential Neighbours Selection in online social networks

EU FP7 RECOGNITION Project

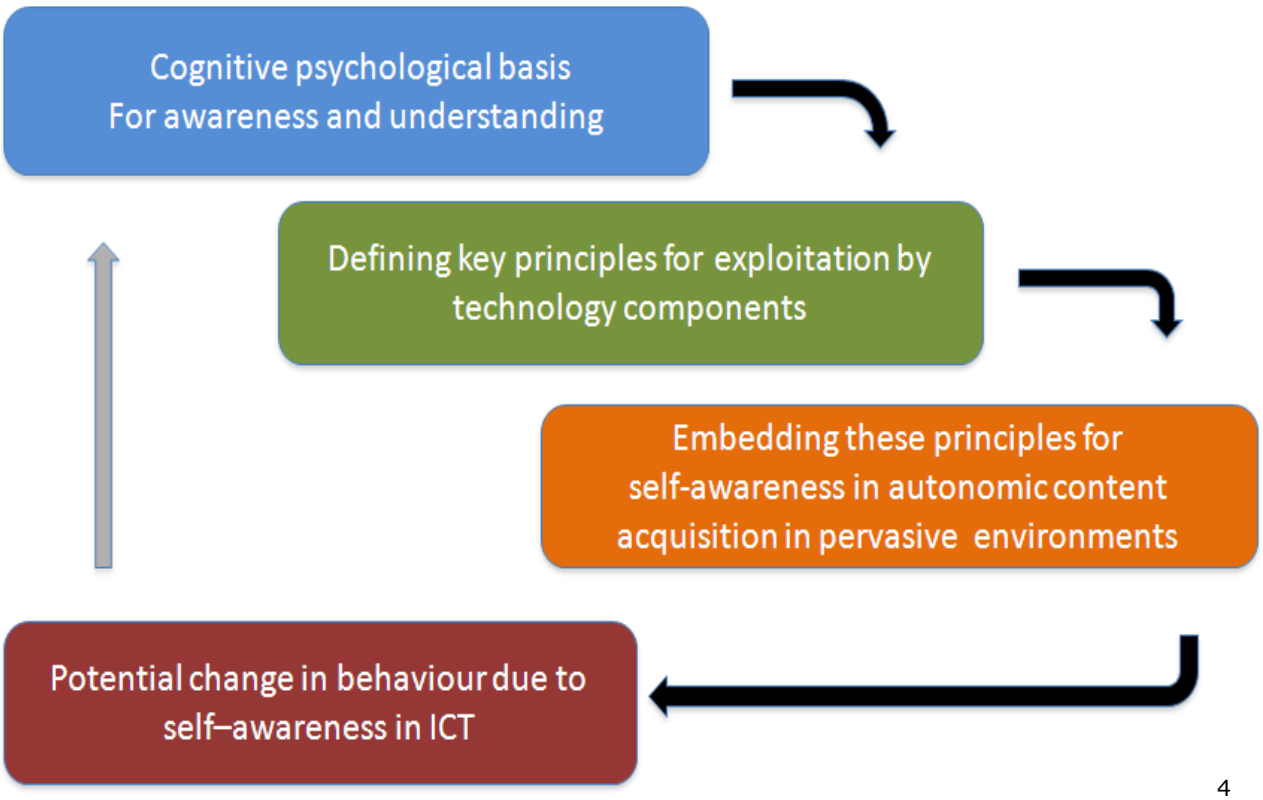


RECOGNITION:
Relevance and Cognition for Self-Awareness in a Content-Centric Internet

- Self-awareness to support ICT function
  - Enabling content centricity
    - Better fitting of users to content and vice versa
  - Synchronise content with human activity and needs
    - Place, time, situation, relevance, context, social search
  - Autonomic management
    - Of content, its acquisition and resource utilisation

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD · UNIVERSITY OF CAMBRIDGE · ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS · EURECOM

Human Awareness Behaviours



Information Propagation in OSN

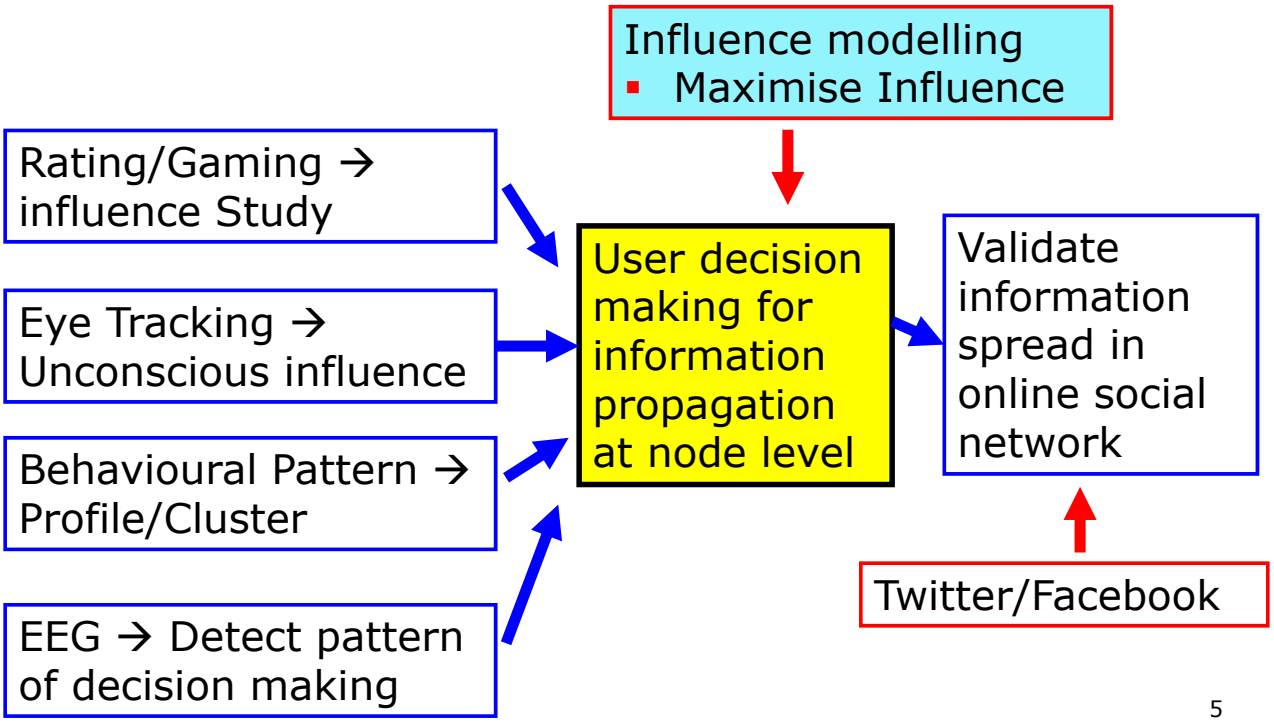
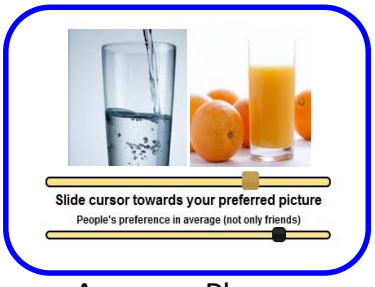


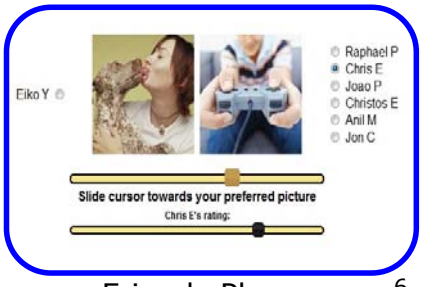
Photo Rating Experiment

- Is the social desirability bias higher when you know the choice of a specific person or the choice is known to friends?
- Is the social desirability bias higher when friends' preferences are displayed instead of anonymous average choices?
- Can we quantify the number of positive reactions, negative reactions and no reactions ?

Blind Phase



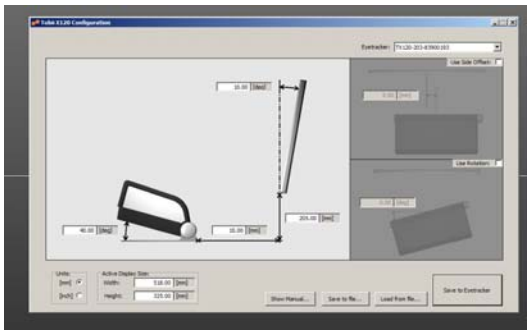
Average Phase



Friends Phase

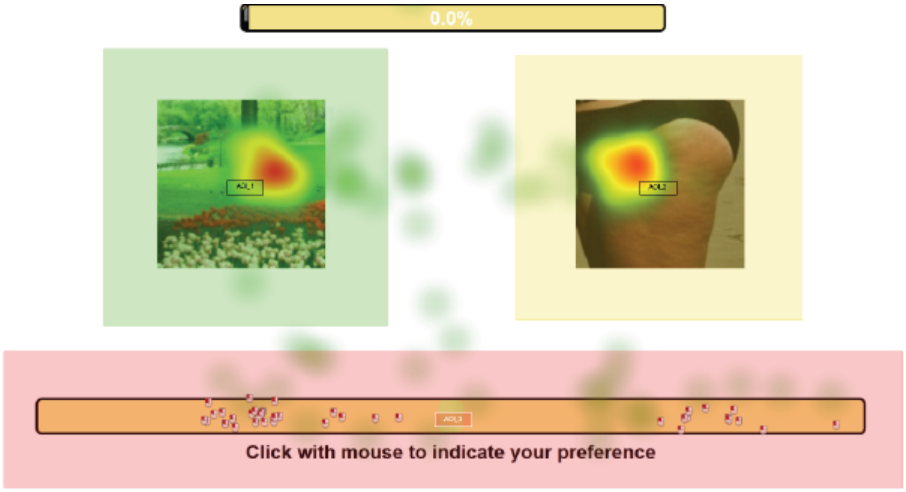
Eye Tracking

- Complement photo rating study by understanding Unconscious Behaviour
- Tobi
  - Only one device
  - Automatically recorded web page coordination



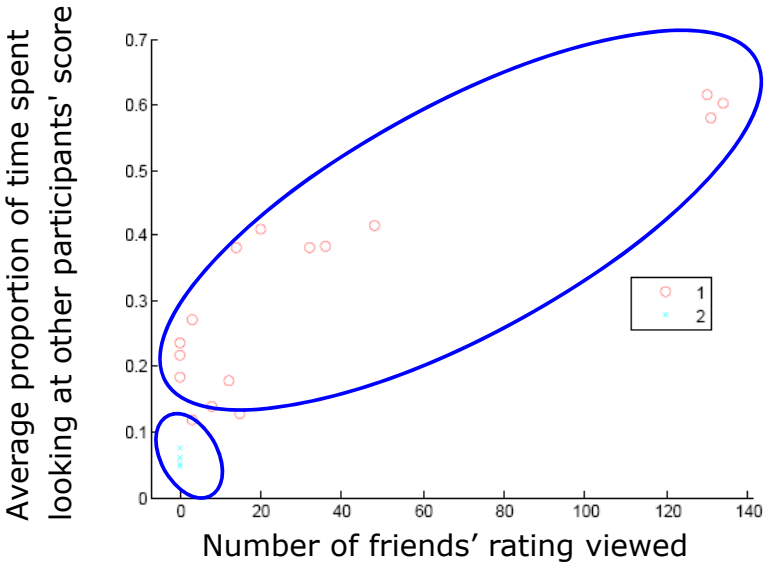
Capture Area of Interest (AOI)

- Using Heat Map tool to view where the Fixations concentrate



Participants Characteristics

- Two groups: Group2 (no click – not looking at others' rates) and Group1 (the other)
- Group2 : changing their opinion significantly less than the others

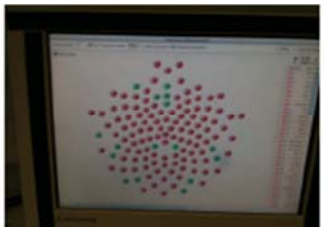


EEG System

- Complement photo rating study and Eye tracking

→ Signal propagation patterns among channels for understanding decision making mechanism

- High-density 128-channel electrolyte-based system (EGI)
- Emotiv Neuroheadset



Word-of-Mouth Effects

90% of consumers trust peer recommendations while only 14% trust advertisement

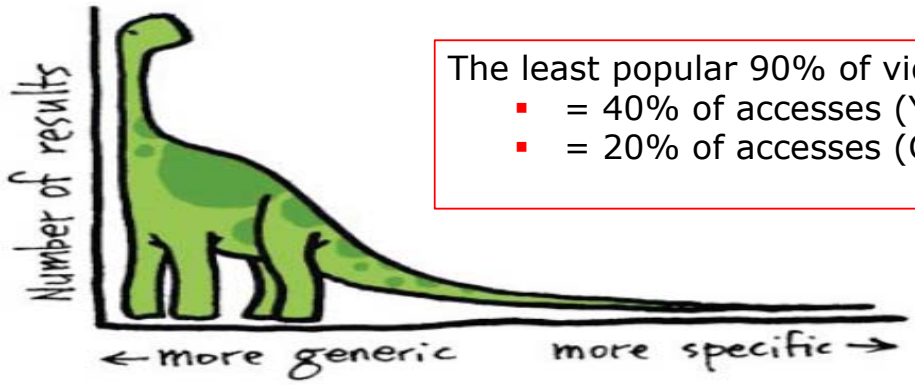
Applications: Prediction of content access, Viral marketing, recommender systems, feed ranking

OSNs for Spread of Influence

- Social network plays a fundamental role as a medium for the spread of influence among its members
  - Opinions, ideas, information, innovation …

Popularity of User Generated Content

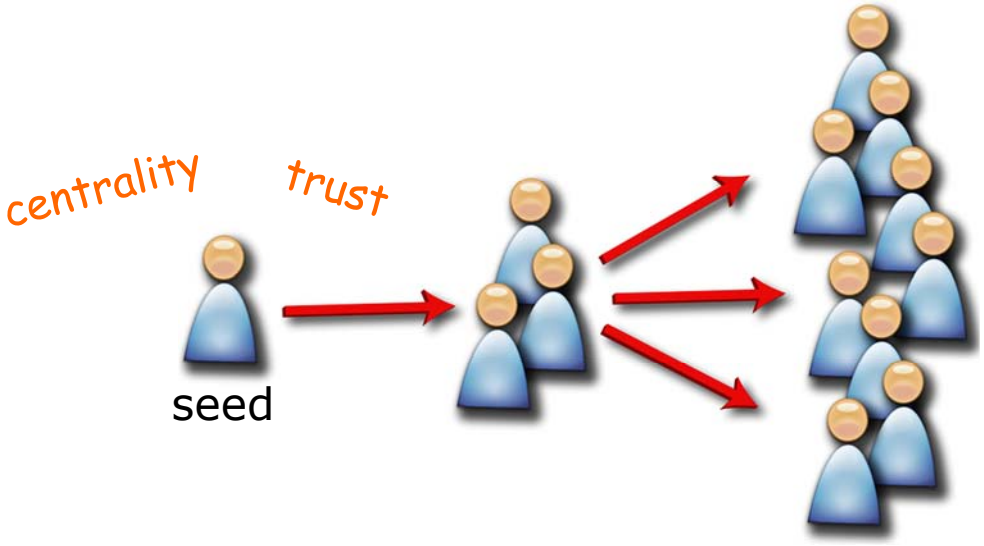
- UGC Access Pattern: Heavy-Tail
  - Small portion of popular content
  - Rare access to tail content – dynamic nature



The least popular 90% of videos
- = 40% of accesses (Yu'06)
- = 20% of accesses (Cha'07)

- Users need to become aware of new rare objects
- A delivery infrastructure is needed

Importance of information seeds

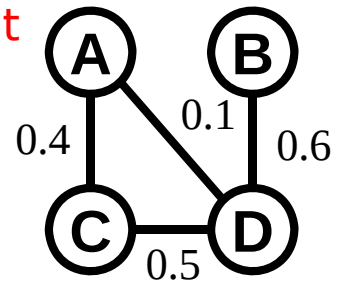


Influence might be changed with information seeds

Influence Maximisation

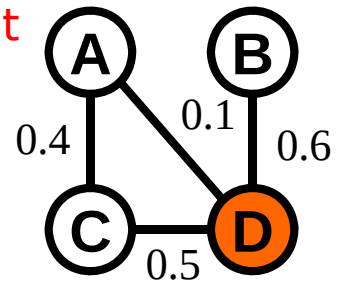
- Problem: (Domingos et al., 2001; Kempe et al., 2003)
  - Given a social graph $G = (V, E)$ with influence probabilities on edges, select k individuals such that by activating them, the expected spread of influence is maximised

Input



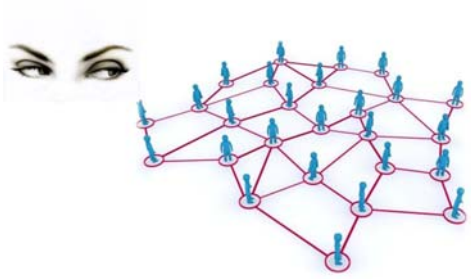
Social graph with influence probabilities of edges

Output



$k = 1$

Limitations of Influence Maximisation



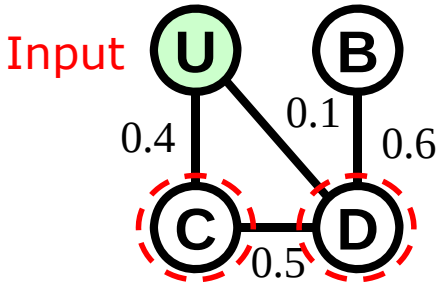
This model requires a bird's eye view of an entire social graph. In real world, who knows the whole network topology?

In practice, a node can initially share the information with only some of its neighbours rather than a set of any arbitrary.

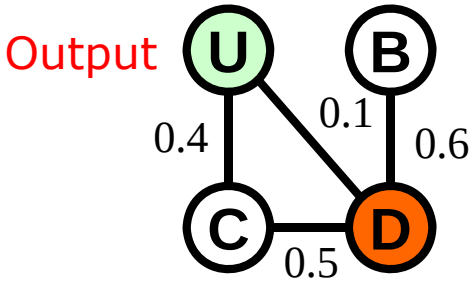
Influential Neighbour Selection

- Influential Neighbours Selection (INS) problem:

Given a social graph $G = (V, E)$ with influence probabilities on edges and a node u, select u's min(k, degree(u)) neighbours such that by activating them, the expected spread of influence is maximised



Social graph with influence probabilities of edges and U



$k = 1$

Research Question

How can the neighbours be effectively chosen for information diffusion in OSNs?

For example, when k=1, we may choose the most powerful(?) neighbour as the activated node

Our Assumptions

1. Each node only communicates with its immediate neighbours
2. Each node has no knowledge about the global network topology
3. Each message size is bounded to O(log |V|) bits
4. For simplification, we use a constant influence probability for all edges

Neighbours Selection Strategies

- Set 4 selection strategies based on local connectivity pattern such as degree and clustering coefficient
  - Random selection
  - High degree selection
  - High volume (proposed by Wehmuth and Ziviani) selection
  - High weighted-volume selection (a good approximation of closeness centrality)

Function	Influence of v	Cost
Ran.	1	$O(1)$
Deg.	$d(v)$	$O(\kappa)$
Vol.	$\sum_{w \in N_h(v)} d(w)$	$O(\kappa^{(h+1)})$
Wei.	$\sum_{w \in N_h(v)} d(w) \cdot (1 - c(w)) \cdot (1/2^{\delta(v,w)})$	$O(\kappa^{(h+1)})$

Datasets for Simulation

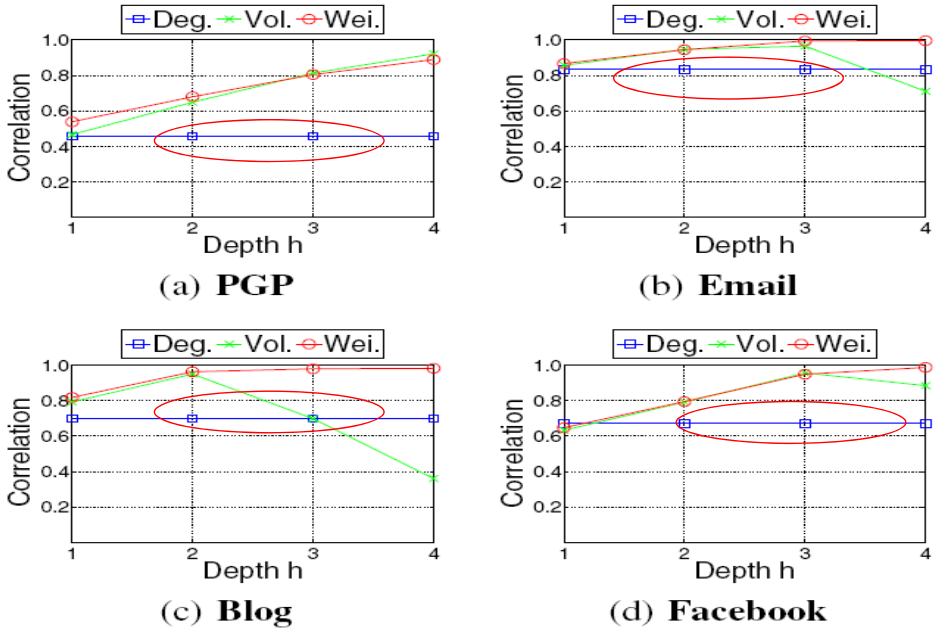
- We test the four real-world network datasets:

Network	$\|V\|$	$\|E\|$	κ	C	D
PGP [6]	10,680	24,316	4.55	1	24
Email [7]	1,134	5,453	9.62	1	8
Blog [8]	1,224	16,718	27.32	2	inf
Facebook	26,701	251,249	18.82	1	15

k: average degree
C: number of connected components
D: network diameter

Correlation coefficients

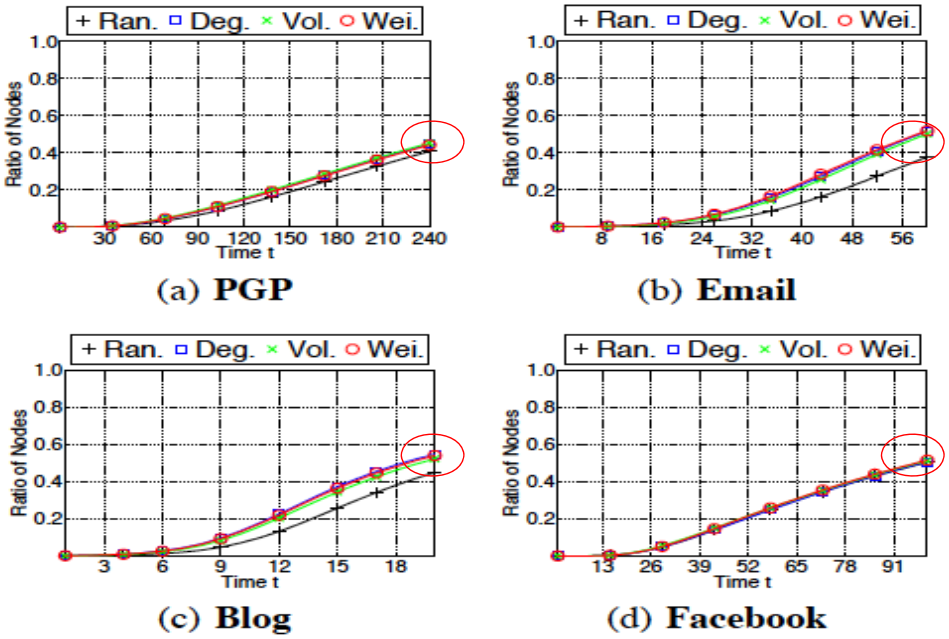
- Pearson correlation coefficients between node property and closeness centrality



h:distance from v

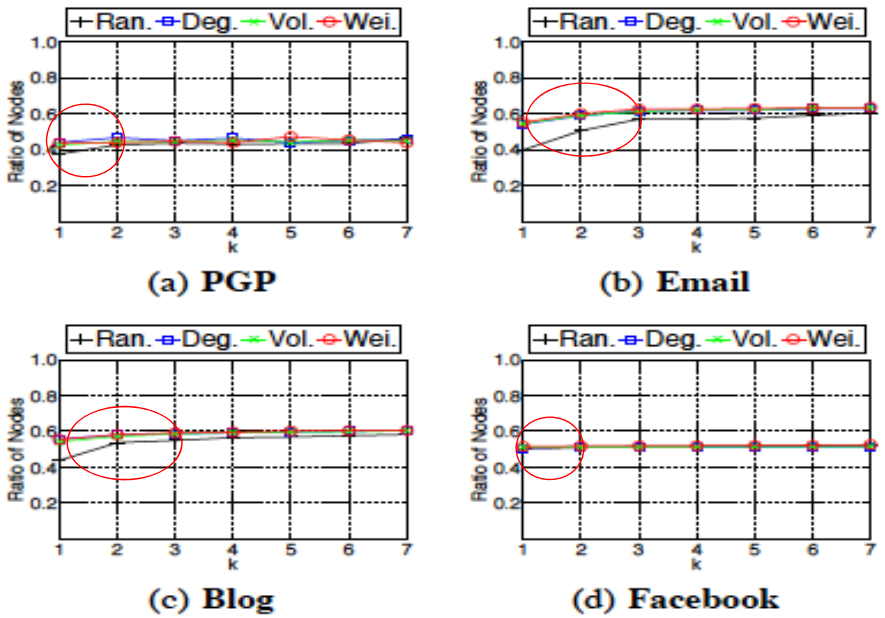
Simulation Results

The ratio of the average number of activated nodes to the total number of nodes in the network over time *t*



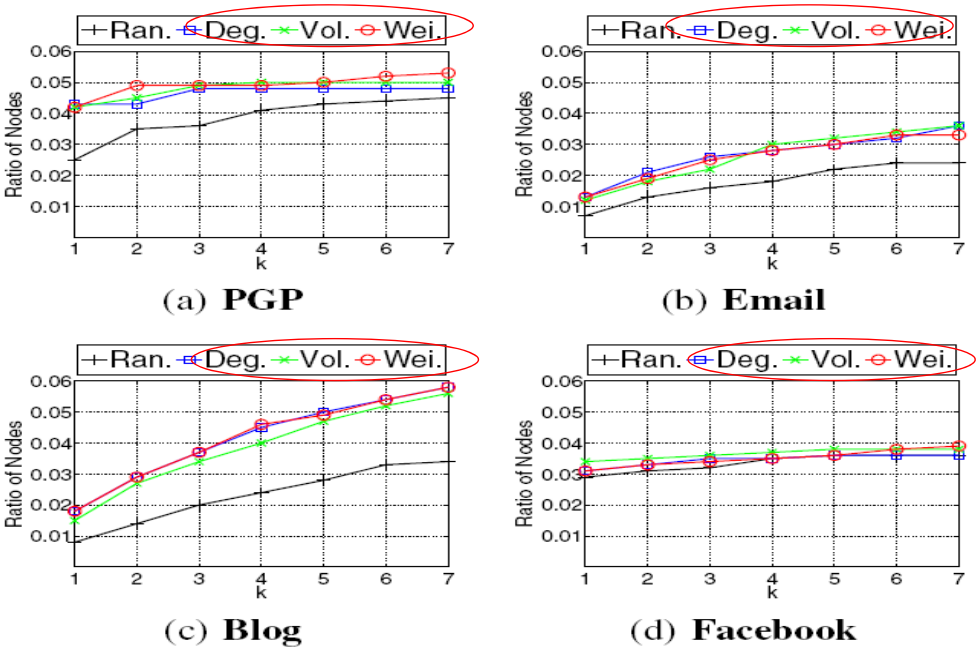
Effect of Size of K – Long Term

The ratio of the average number of activated nodes to the total number of nodes with the number of initial activated neighbours *k*



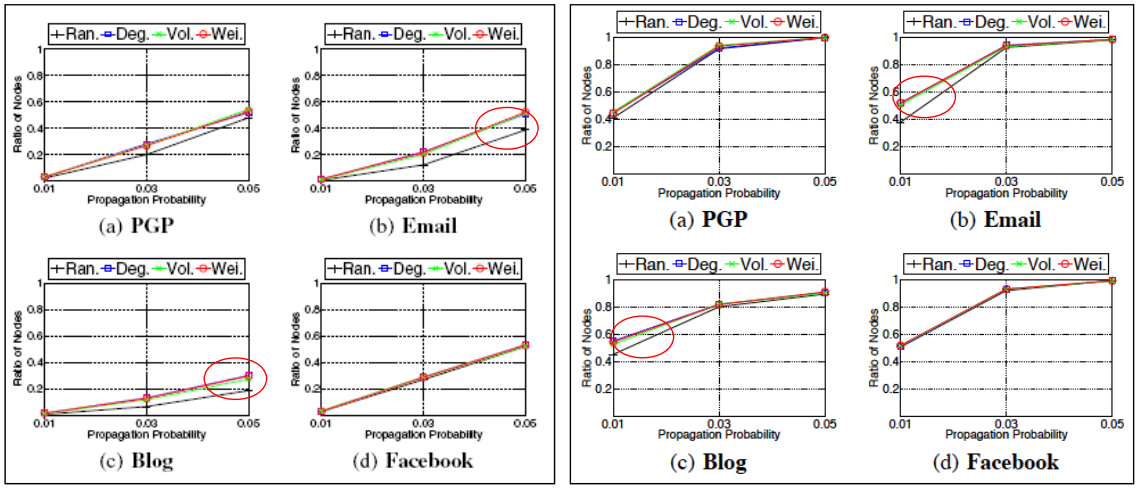
Impact of Size of K – Short Term

Changes in ratio of the average number of activated nodes to total number of nodes with the number of initial activated neighbours k (1/4 of full timeline)



Propagation Probability

The ratio of the average number of activated nodes to the total number of nodes with influence probability λ (k=1)



Short Term

increase gap

Long Term

decrease gap

Conclusions

- We introduce a new problem called INS problem to select a node's neighbours to efficiently disseminate its information
- We empirically test the four reasonable selection strategies through intensive simulation based on four real-world network topologies
  - We recommend using the degree selection strategy for short-term propagation but the random selection strategy for long-term propagation to cover more than half of a network, respectively
  - Volume and Weighted produce similar results to those obtained by degree - we recommend using degree, which less costs

Future Work

- We plan to test community based selection methods
  - If a user's neighbours are divided into several disjoint communities, we may improve the performance of information diffusion by selecting initially activated neighbours from different groups, respectively
- We develop a more general model for information diffusion. We may consider not only a user's neighbours but also neighbours of neighbours as the candidate space of the initially activated nodes
- Consideration of activation delay

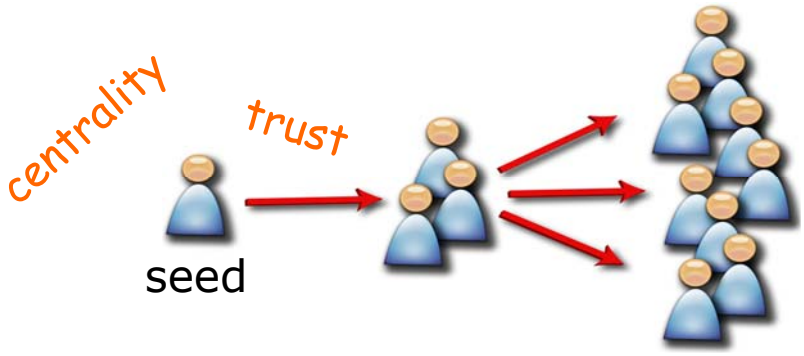
Propagation Decision at Node

- Individual influence probability λ.
- Constant probability value is used on the decision making if the information would get propagated or not at each node

→ need to model decision making mechanism at each node

→ Psychological Behaviour Embedded Model

Questions?



Paper: Influential Neighbours Selection for Information Diffusion in Online Social Networks in 2012 ICCCN

http://www.cl.cam.ac.uk/~ey204/pubs/2012_ICCCN.pdf