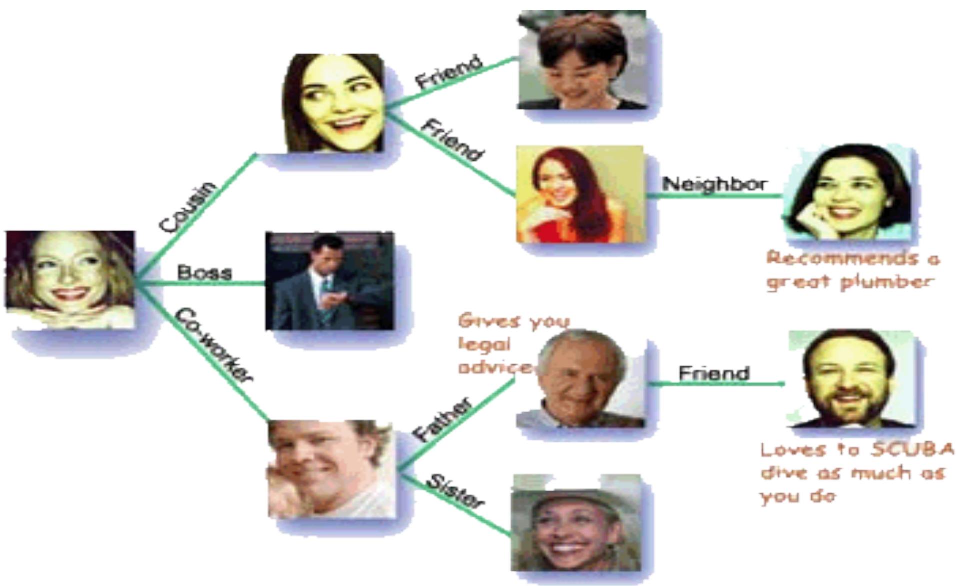
Accelerating Problem-Solving in Collaborative Social Networks

Yi Chen (Arizona State Univ)
Shu Tao (IBM Research, USA)
Xifeng Yan (Univ. of California, Santa Barbara)
Presenter: Ziyang Liu (Arizona State Univ)

Thanks to Gengxin Miao, Louise Moser, Nikos Anerousis and National Science Foundation

Motivation

- Problem solving is ubiquitous in everyday life, enterprise, governments, etc



Problem-solving in daily life

Customer service

Aardvark

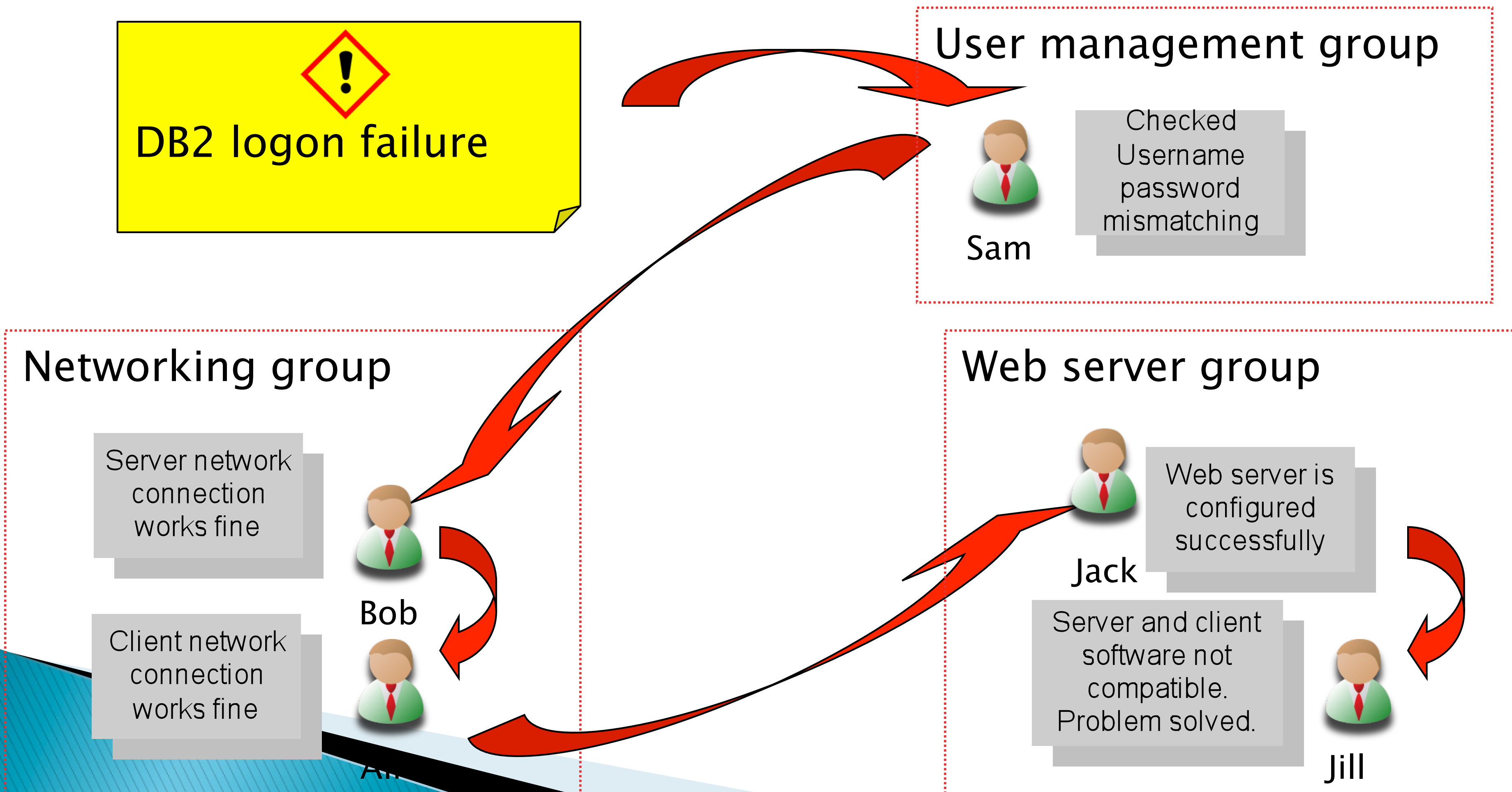
Social search engine

- It’s challenging to find problem resolver (even with the assumption that each problem has a single resolver)
  - Great diversity of the problems
  - Great diversity of people’s expertise

Problem Routing

- Often a problem is routed among multiple experts in a collaborative social network before it reaches a resolver

➡ Problem-solving workflows



Problem Definition

A set of problems reported

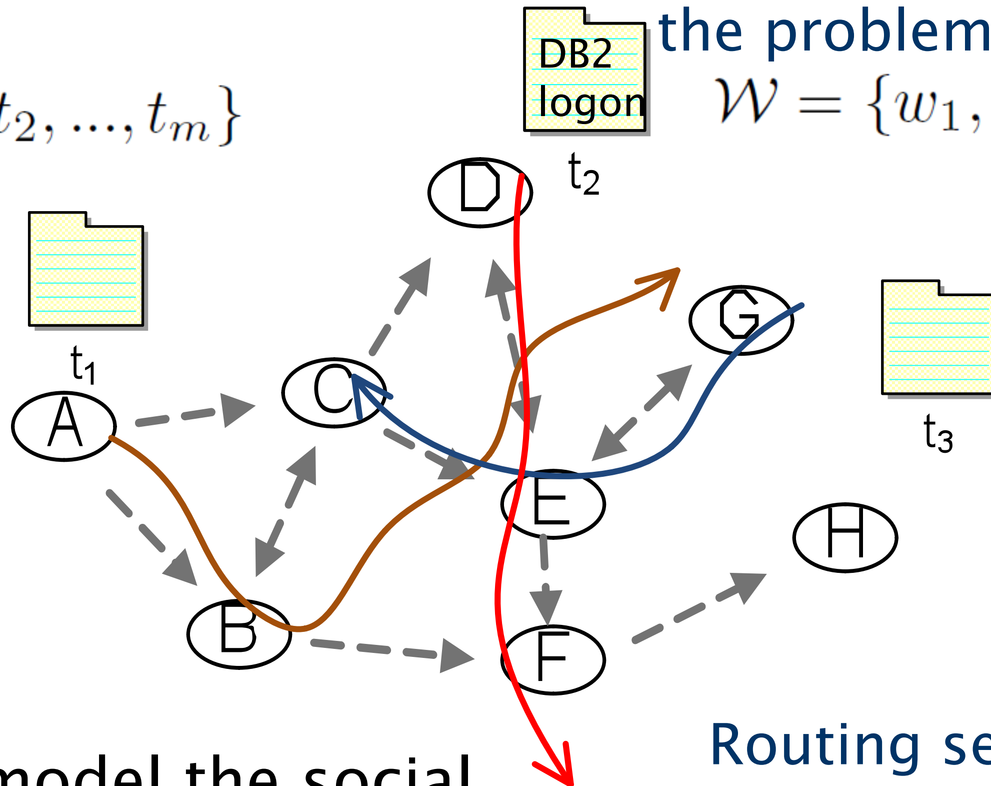
$\mathcal{T} = \{t_1, t_2, ..., t_m\}$

Word description of the problems

$\mathcal{W} = \{w_1, w_2, ..., w_n\}$

An interconnected network of experts

$\mathcal{G} = \{g_1, g_2, ..., g_L\}$



Routing sequence of the problems

$R(t) = g_{init}(t) -> ... -> g_{res}(t)$

Challenges:

- How to mine and model the social network?
- How to expedite social search?
- How to quantify collaboration effectiveness?

A social network,
Workflows of problem-solving
and their interaction

I. Modeling the Social Network – A Graph Model

- First attempt
  - Experts are represented as nodes
  - Interaction between the experts are reflected in the routing decisions.

    an edge between nodes if there exists a problem routed between them

- This model does not capture the rich interactions between experts

- Challenges:
  - How to capture the collaboration patterns between experts?

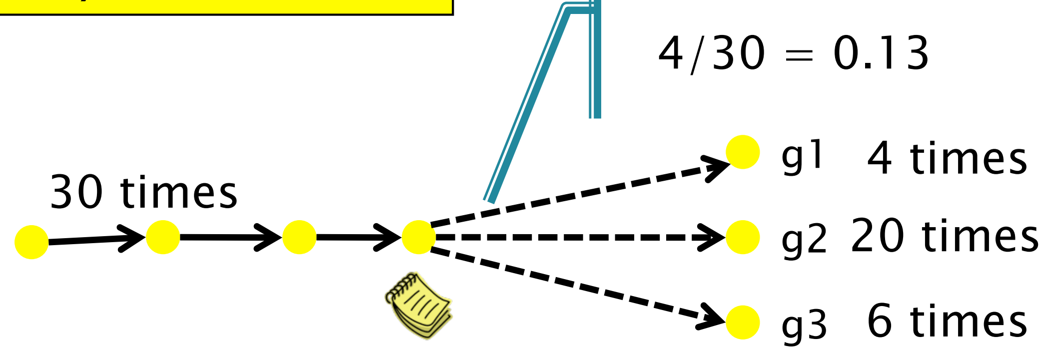
I. Modeling the Social Network – Markov Models

- We use a Markov model to capture the high-order interaction between experts.
- Transition probabilities: given a set of previous experts, the probability of the problem being transferred to expert gi, as the next step

$$P(g_i \mid S_{(k)}) = \begin{cases} N(g_i, S_{(k)}) / N(S_{(k)}) & \textit{if} \quad N(S_{(k)}) \neq 0 \\ 0 & \textit{otherwise} \end{cases}$$

# of instances that a problem is transferred to group *gi , after being processed by S(k)*

# of instances with a set of group transfers



I. Modeling the Social Network – Determine the Markov Order

- The higher order, the better prediction accuracy

  Use conditional entropy to measure predictability

$$H(g \mid S_{(k)}) = -\sum_{S(k)\in g_k} P(S_{(k)}) \sum_{g\in G} P(g \mid S_{(k)}) \log P(g \mid S_{(k)})$$

- The higher the order, the more complex the system
- Find a right tradeoff for order k:

$$H(g \mid S_{(k)}) - H(g \mid S_{(k+1)}) < \theta$$

  Beyond threshold, the improvement of predictability is small

- A higher order may not be feasible due to the limited size of historical data, use a lower order when necessary.

II. Expediting Social Search: – Intuition

- Challenge: given a set of historical problems and their resolution workflows, how to recommend the routing of a new problem?

- Observations:
  - The majority of historic routing decisions are correct
  - A mis-routing decision may cause cascading effects and result in long routing sequence

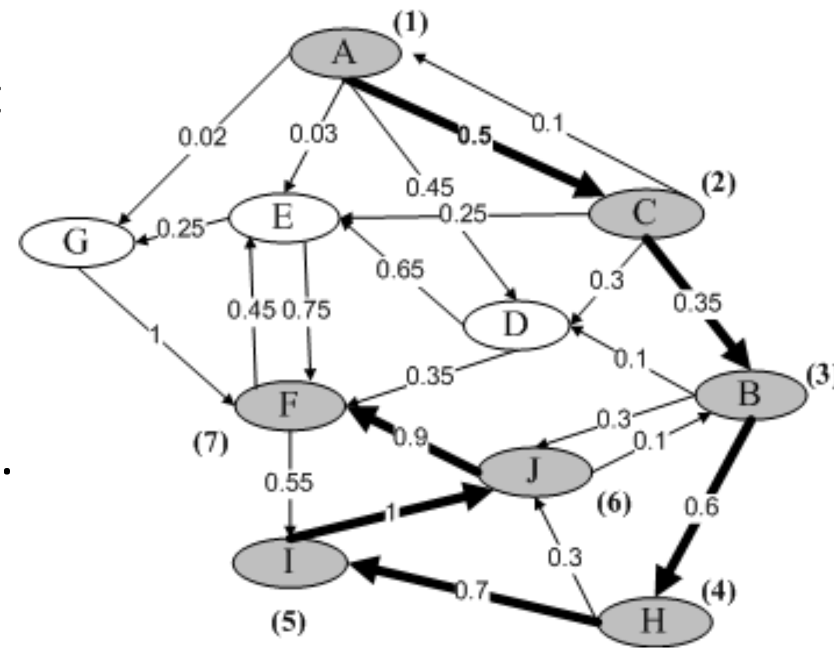
- Intuitions: use the patterns captured in the social network model to guide future problem routing

II. Expediting Social Search: – Graph Search

- Consider the first order Markov model, and do a depth-first search on the Markov graph
  - Choose the next node with the highest transition probability.

    $$g^* = \arg\max{}_g P(g \mid g_{(l)}, \forall g \in G)$$

  - Stop until find the resolver or reach a node without unvisited neighbors
  - Same node should not be visited twice.



- Drawbacks
  - Only considers first order, which may not be good
  - Only relies on the <u>current node</u> to make transfer decisions. Thus if current node is a result of mis-routing, the mistake may be propagated.

9

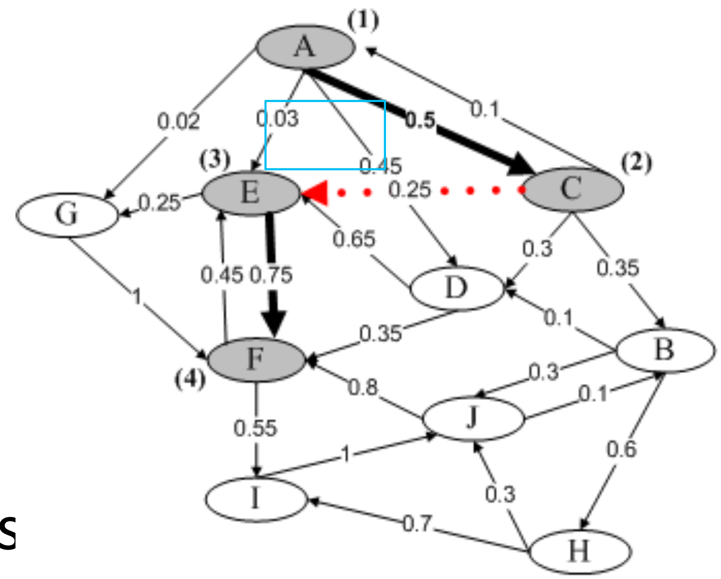
II. Expediting Social Search: – Maximize Likelihood

- Consider variable length of historical routing patterns
  - Check all available transfer probabilities for all Markov orders
- Consider multiple active state:
  - Consider every node visited in the past as a candidate for seeking next node
- Select the next node g that gives maximal transfer probability

$$g^* = \arg\max{}_g P(g \mid g_{(l)}), \forall g \in L_c, S_{(k)} \subseteq L_v$$

Lv: visited node set

Lc: candidate node set



(a) Routing Steps in the model

$S_{(2)}$	Next group g	$P(g \mid S_{(2)})$
{A,C}	E	0.6
{A,C}	D	0.2
{A,C}	B	0.1
{A,C}	G	0.1
{E,C}	F	0.6
{E,C}	G	0.4
...	...	...

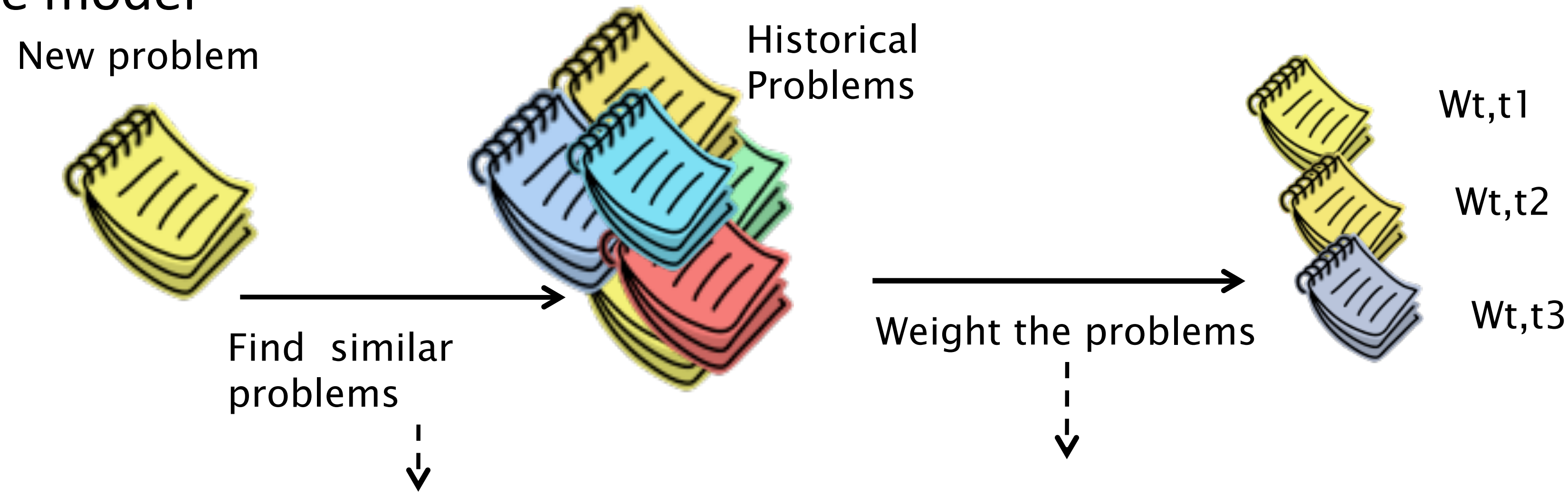
(b) 2nd order probability table

Why not ALWAYS higher order?

1) Unavailable due to the limited historic data
2) Lower order patterns may have a higher confidence

II. Expediting Social Search: – Context Aware Routing

- The first two approaches consider the routing sequences of historical problems only.
- Context-aware routing: leverage context provided in problem content
- Historic problems that are similar to the current one are used to train the model



$$Cosine\ Similarity, Cos(V, Vi) = \frac{V \, . \, Vi}{|V| \, . \, |Vi|}$$

$$Wt, ti = \cos(V, Vi`)^{m}$$

Where, V and Vi` are vectors derived from problems T and Ti`

We use a exponential parameter m to tune the weigh for content similarity.

III. Quantify Collaboration Effectiveness

- Two Key factors that affect the efficiency of problem solving
  - Technical expertise

    Determines how well/efficient that a problem can be solved by an expert

  - Awareness of technical expertise of other groups.

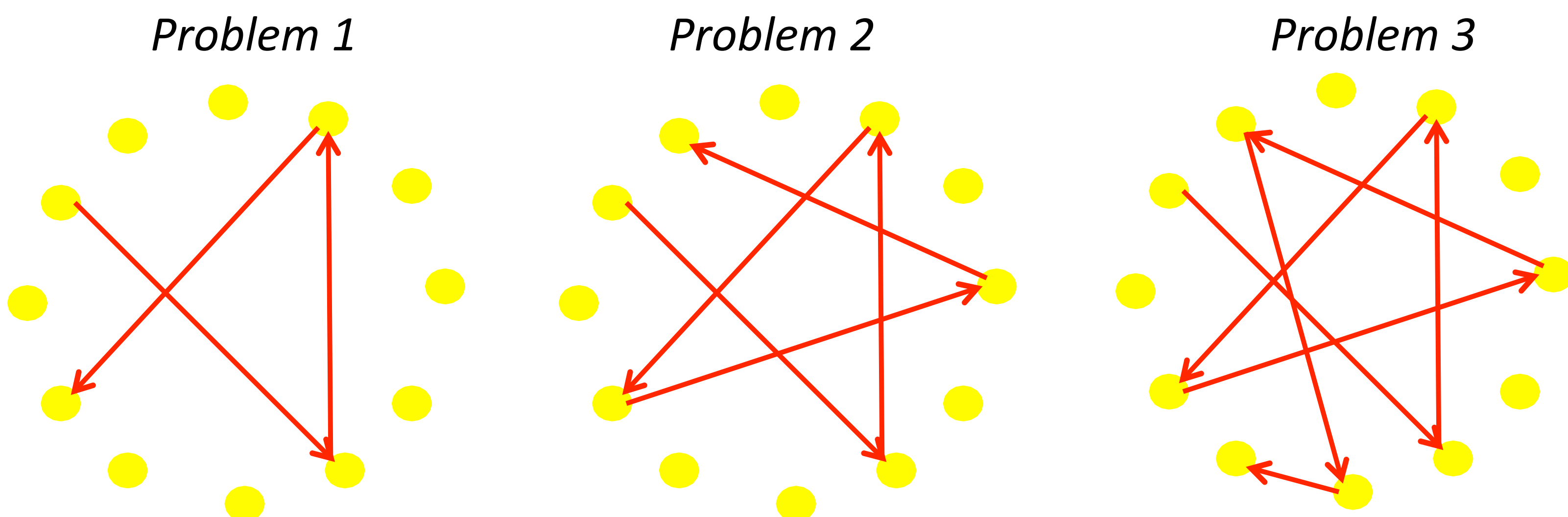
    Determines the effectiveness of collaboration, and thus the efficiency of a problem to be routed to the right expert

- **Challenge:** how to evaluate collaboration effectiveness?
  - Traditional approaches are based surveys, which are often tedious, time-consuming and error prone.
  - Is it possible for a quantitative assessment in a computational framework?

III. Quantify Collaboration Effectiveness -2

- Key idea: we evaluate the performance of a node by considering the performance of the network with versus without the node.
- How to measure the performance of the network?

  Mean Number of Steps to Resolve (MSTR)



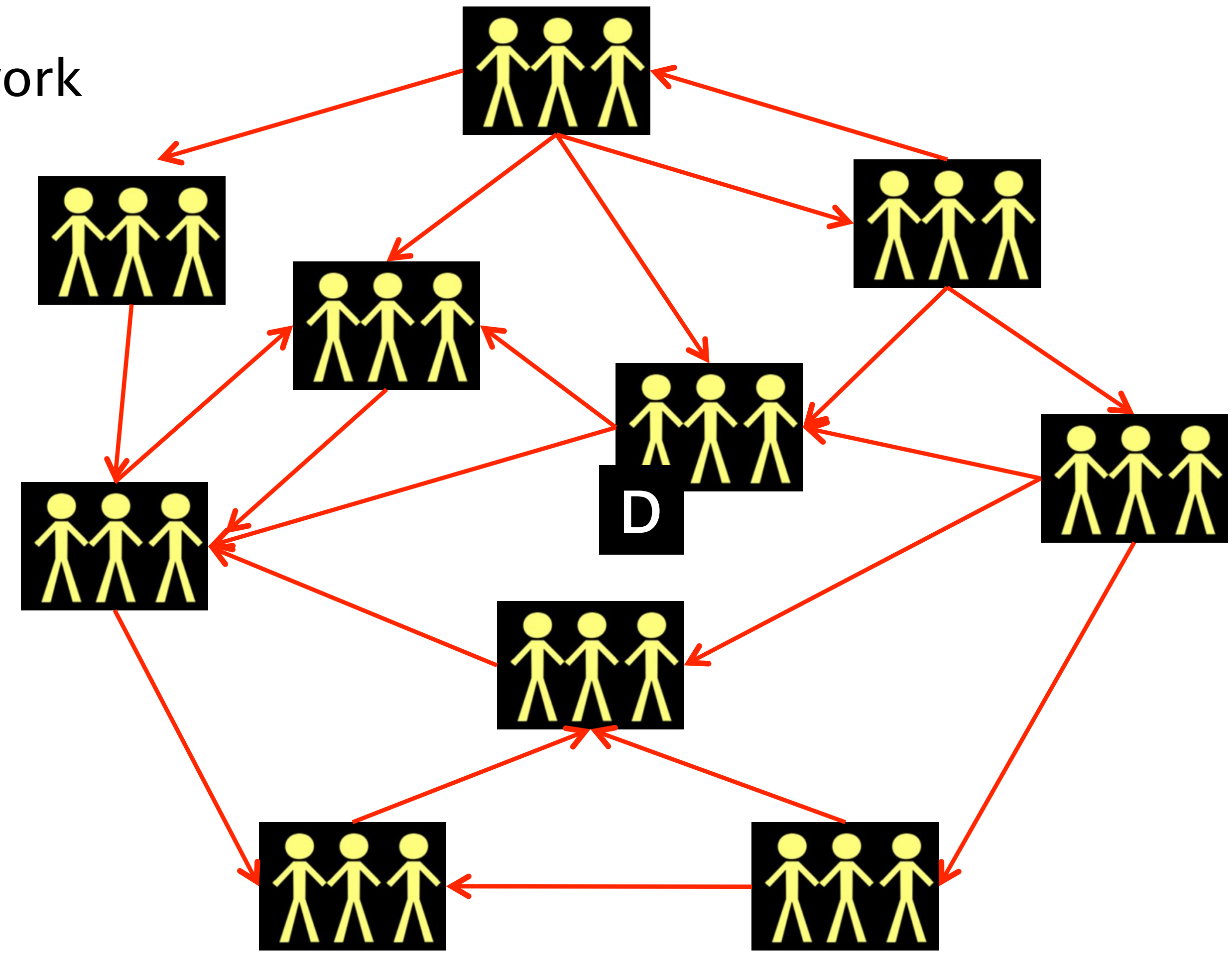
No. Steps to resolve = 3 No. Steps to resolve = 5 No. Steps to resolve = 7

MSTR = Total number of steps to resolve for m tickets / m = (3+ 5 + 7) /3 = 5

III. Quantify Collaboration Effectiveness – Illustration

Problem Resolution Network without D : M`

Problem Resolution Network : M



III. Quantify Collaboration Effectiveness – Illustration (cont)

A network M with node D

Problem	No. of Steps to resolve
1	14
2	18
3	24
…	…
…	…
m	25

A network M' without node D

Ticket #	No. of Steps to resolve
1	6
2	8
3	10
…	…
…	…
n	11

- We compare MSTR(M) and MSTR(M') to judge the effectiveness of node D as a collaborator
- To determine whether network M` has significant difference with network M, we adopted standard **T-test.**

III. Quantify Collaboration Effectiveness –Simulation

- How to obtain the performance of the network after excluding a node?
- Removing a node from a real network is infeasible.
- The need of simulating the network and the problem-solving workflows

- The social network model and the routing recommendation engine discussed so far captures the majority decisions in the history.

  Thus, they provide a simulation of the network and workflows.

Some Experimental Evaluation

- Data set: problem tickets from IBM's problem management system over a 1-year period from Jan 1, 2006 to Dec 31, 2006.
- These tickets were classified into 553 problem categories, e.g., AIX, DB2, Windows etc.
- On an average 50-1000 groups were involved in solving tickets of each problem category.
- The data is divided into training set and testing set

Improvement on Routing efficiency

- Improvement of different problem categories on MSTR (Mean Number of Steps to Resolve)

Category	Original	VMS	MSTR (% off)
ADSM	5.37	3.23	37.99%
AIX	4.89	2.78	43.15%
BIOS	4.49	2.94	34.52%
DB2	4.78	2.57	46.23%
WINDOWS	3.93	2.86	27.23%
All Categories	3.94	2.58	34.52%

Human decision Our work Improvement

18

Evaluation of Collaboration Effectiveness

- The groups that are identified as ineffective collaborators in our evaluation make unnecessary transfers in case studies.

Entry	Description
New Ticket 209366	Not showing image at all on eSMRT. We had some errors on the image
Transferred to Group MRTOP	Problem status has been updated to Open
Transferred to Group MRTPR	ETL deliver data..
Transferred to Group MRTEX	Transfer to MRTDB
Transferred to Group MRTDB	The metric table is blank for ID 932, thus the SL is not showing up on the topshet. Resolution: modified roll-up code to have metric 932 shown up in metric.

Unnecessary transfer to MRTEX

Conclusions & Future Work

- We proposed an innovative model for collaborative social networks and their problem-solving workflows

- Techniques for routing recommendations are developed

- A quantitative framework for evaluating collaboration effectiveness is proposed

- We plan to tackle the challenges with relaxing the assumptions of a single resolver of each problem and unicast.

Thank you!

Questions?