Analyzing Expert Behaviors in Collaborative Networks

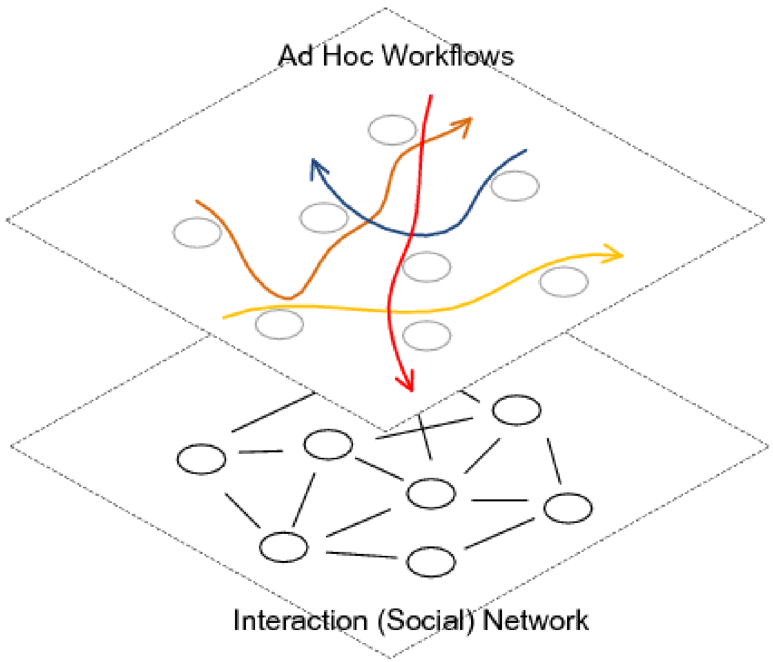
Huan Sun

University of California, Santa Barbara

KDD 2014, Aug 26th

In collaboration with Mudhakar Srivatsa (IBM Research), Shulong Tan (UCSB), Yang Li (UCSB), Lance M. Kaplan(Army Research Lab), Shu Tao (IBM), Xifeng Yan (UCSB)

Collaborative Networks



Information flows

Enterprise ticket

ID	Day	Entry
81	05-14	New Ticket: DB2 login fail
81	05-14	Transfer to Group SMRDX
81	05-14	Contacted Mary for recycling
81	05-14	Transfer to Group SSDSISAP
81	05-14	Status updated ...
81	05-15	Transfer to Group ASWWCUST
81	05-15	Web service checking
81	05-18	Could not solve the problem.
...	...	...
81	05-18	Transfer to Group SSSAPHWOA
81	05-22	Resolved

Eclipse bug record

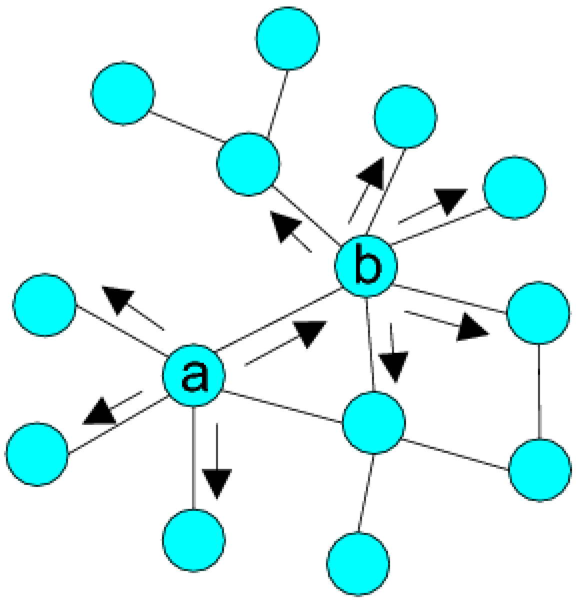
Bug description:
NullPointerException referencing non-existing plugins.

Who	When	Description
dean	2001-11-01 07:17:38 EST	Added component Core. Reassigned.
rodrigo	2001-11-20 18:53:40 EST	Added component UI. Reassigned.
dejan	2002-01-09 20:46:27 EST	Converted the unresolved plugin to a link. Fixed.

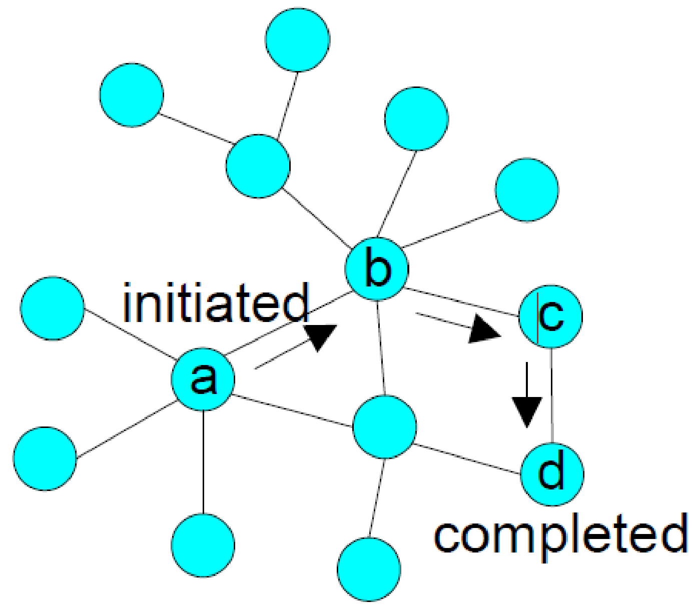
https://bugs.eclipse.org/bugs/show_activity.cgi?id=325

Differences: Social Networks VS. Collaborative Networks

Social network
- information diffusion



Collaborative network
- information routing



Core Problems

- **How do experts make routing decisions?**
- **Who have made inefficient routing decisions?**
- **How to optimize the routing performance through targeted training?**
- **Can the completion time of a task be predicted so that one can act early for difficult tasks?**

Observations on Routing Decision Making

Observation 1

Tasks with similar content, but different routing sequences

e.g., two problem tickets in IBM IT service department

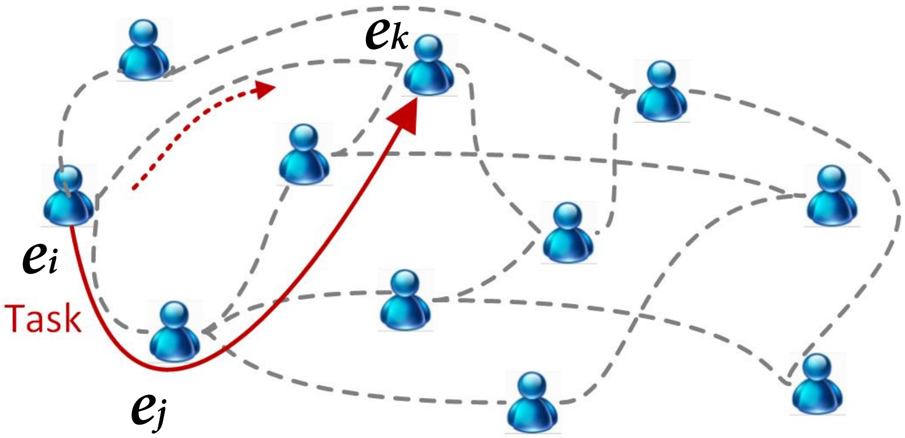
Task ID	Task Content	Routing Sequence
492	Need password reset for kasperj on machine "pathfinder"". Route to NUS_N_DCRCHAIX	12 →505 → 1914→ 1915 →1916 →247
494	Need password reset for jhallacy on machine ""pathfinder"". Route to NUS_N_DCRCHAIX	12 → 13 → 86

Routing decision is not deterministic, given a certain task.

Observations on Routing Decision Making

Observation 2

An expert might not directly send a task to a resolver.



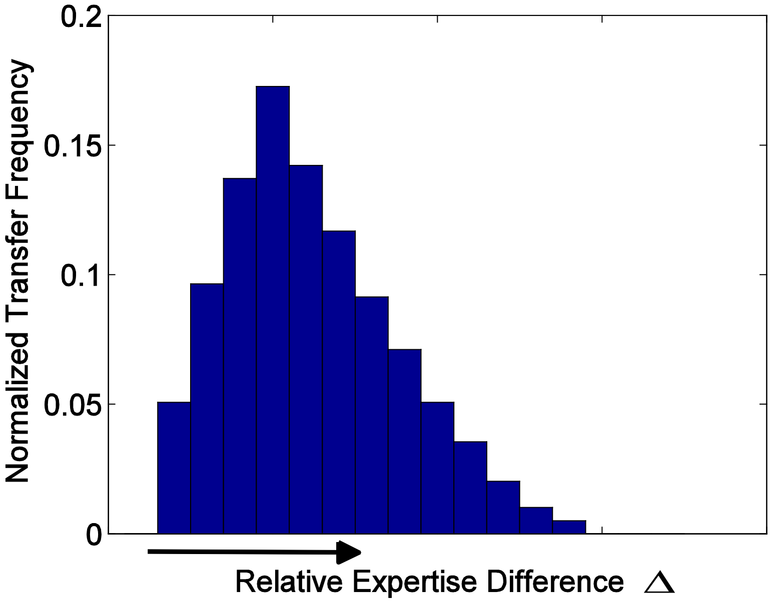
Is it because he does not understand the task very well, thus randomly routing it? Or

he believes the other expert has a better chance to solve it, or a better chance to find the right expert to solve it?

Observations on Routing Decision Making

Observation 3

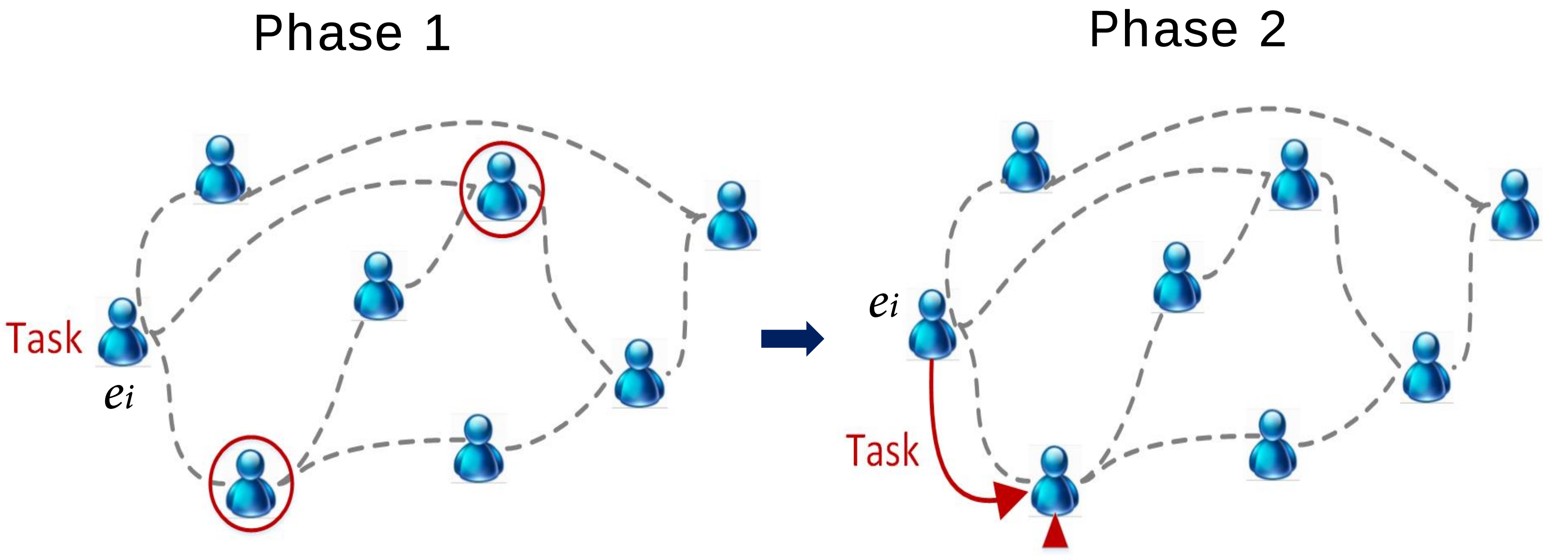
An expert tends to transfer a task to some expert whose expertise is *neither* too close *nor* too far from his own. That is, not necessarily the final resolver!



Modeling Expert Decision Logic

- A Two-Phase Assumption

When an expert transfers a task,

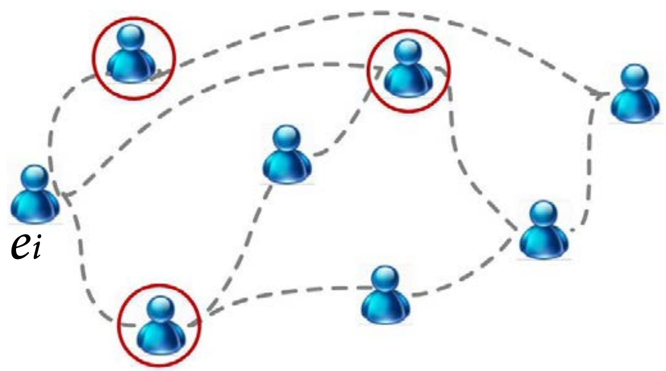
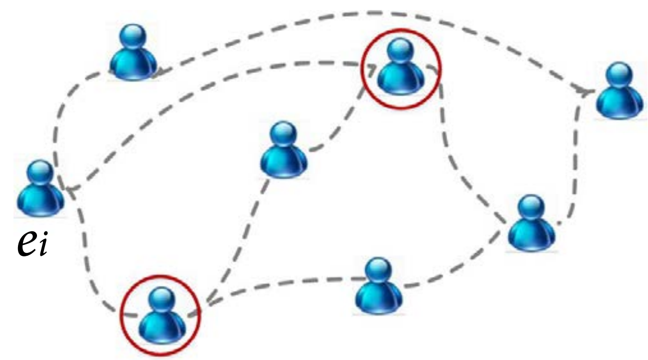


Candidate pool establishment

Specific expert selection

Modeling Expert Decision Logic

- Phase 1: The establishment of candidate pool C.

Two Routing Strategies	
Task-Neutral Routing (TNR)	**Task-Specific Routing (TSR)**
C: All the neighbors (all the experts he has contacted)	C: experts in one 's neighborhood, but estimated capable of solving the task
	

Modeling Expert Decision Logic

- Phase 1: How to decide C in Task-Specific Routing?

  -- Estimate expert knowledge and capability based on the tasks he has dealt with before.

  -- Logistic model

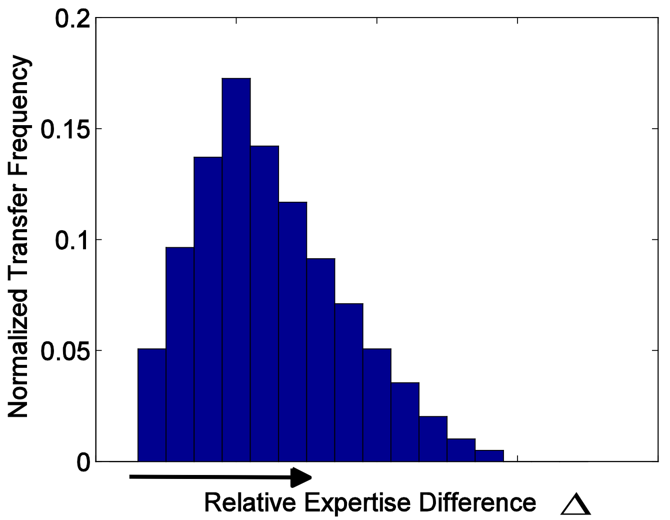
$$\boxed{P(e_i, t) = \frac{1}{1 + \exp(-(W_1 t + W_2 e_i + b))}}$$

Modeling Expert Decision Logic

- Phase 2: Which expert in C to transfer the task to ?
  - Uniform Random
  - Volume-biased Random
    - Volume v_{ij} : the number of tasks dispatched from e_i to e_j

$$\boxed{P(e_i \xrightarrow{t} e_j) \propto v_{ij}}$$

  - Expertise Difference



log-normal density:

$$f \propto \frac{1}{\Delta} e^{-\frac{[\ln \Delta - \mu]^2}{2\sigma^2}}$$

$$\boxed{P(e_i \xrightarrow{t} e_j) \propto f(\Delta(e_i, e_j))}$$

Modeling Expert Decision Logic

- 6 particular routing patterns

	Task-Neutral Routing (TNR)	Task-Specific Routing (TSR)
Uniform Random (UR)	TNR^{ur}	TSR^{ur}
Volume-biased Random (VR)	TNR^{vr}	TSR^{vr}
Expertise Difference (EX)	TNR^{ex}	TSR^{ex}

Modeling Expert Decision Logic

- Generative model

Assume: the routing decision of an expert is generated through a mixture of 6 routing patterns.

Decision generation process:

For each expert e_i to transfer tasks,

-Draw the mixture weights of 6 routing patterns:

$$\theta_i \sim \text{Dirichlet}(\alpha)$$

(reflecting e_i 's preferences over adopting different routing patterns)

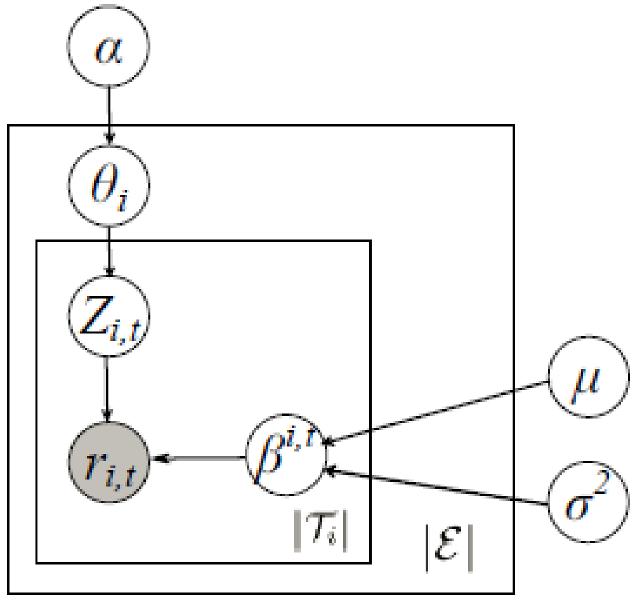
-For each task t to be transferred by expert e_i,

* Draw a pattern label: $Z_{i,t} \sim \text{Mult}(\theta_i)$

* Draw an expert from the candidate pool to receive t.

Modeling Expert Decision Logic

- Generative model



Graphical representation

The likelihood of observing all the task transfer relationships as:

$$\boxed{\mathcal{L} = P(e_i \xrightarrow{t} r_{i,t}, \forall t \in \mathcal{T}_i, \forall e_i \in \mathcal{E} | \alpha, \mu, \sigma^2)}$$

Parameters are optimized by

$$\boxed{\arg\max_{\alpha,\mu,\sigma^2} \log \mathcal{L}}$$

through efficient variational EM algorithm.

Experiments

- Three Datasets

  Real-world problem ticket data collected from a problem ticketing system, in an IBM IT service department.

Datasets	Description	# of tasks	# of experts	% of tasks with CT		
				=2	=3	>=4
DB2	Database usage	26,740	55	44.2	34.3	21.5
WebSphere	Enterprise software	65,786	234	39.0	36.2	24.8
AIX	Operating system	120,780	404	40.0	39.4	20.6

Experiments

- Evaluation Measure
  - Completion Time (CT): Number of experts contacted before a task is resolved
  - Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{|\text{Test Set}|} \sum_{t \in \text{Test Set}} |\widehat{\text{CT}}_t - \text{CT}_t|$$

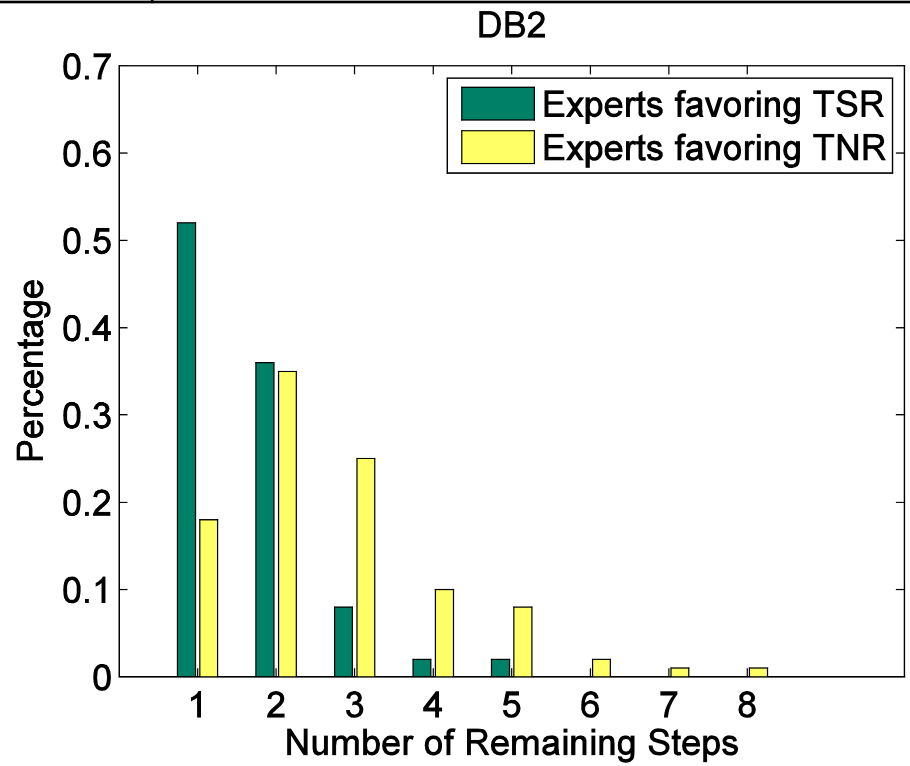
- Results

DB2	
Models	**MAE**
TNR+TSR	**0.08**
Miao *et al*.	0.68
Support Vector Regression	0.80
Bayesian regression	0.84

Experiments

- Resolution Efficiency of TNR vs. TSR

Experts favoring TNR	Weight on TNR patterns $>$ Weight on TSR patterns.
Experts favoring TSR	Weight on TNR patterns $<$ Weight on TSR patterns



Application: Optimizing Collaborations

- Which expert should be trained first to adopt TSR? How much efficiency improvement can we expect?
  - Random: random selection
  - Frequent transferor: select the expert who transfers the most tasks
  - Least efficient: select the least efficient expert

Methods	**Efficiency Improvement (%)**
Random	0.27
Frequent Transferor	0.91
Least Efficient	1.21
Recommendation using our model	2.75

Conclusions

Our model can address the following questions

- **How do experts make routing decisions? (partially)**
- **Who have made inefficient routing decisions?**
- **How to optimize the routing performance through targeted training?**
- **Can the completion time of a task be predicted so that one can act early for difficult tasks?**

Thank you!