



DISCUS

Mining Social Networks in Message Boards

Naohiro Matsumura (Osaka Univ.)

David E. Goldberg (UIUC)

Xavier Llorà (UIUC)

What're you doing here? 4 Answers

- Gusz Eiben: Have you gone soft?
 - Glad GAs are “hard.”
- April 15 is tax deadline.
- Nao Matsumura stuck at Osaka University.
- DISCUS: Distributed innovation & scalable collaboration in uncertain settings.

Four-Quads of HC Interaction

Recombination agent	<i>computational</i>	Standard Genetic Algorithms	Interactive Genetic Algorithms
	<i>human</i>	Computer Aided Design (CAD)	Human Based Genetic Algorithms
		<i>computational</i>	<i>human</i>

Selection agent

Introduction (1/2)

- Well-planned communication realizes **human innovation** in communication [Goldberg 2002]
- **A social network** shows the relationships between individuals in a group or organization where we can observe their social activities.

Ex.)

- Individuals who share a serious problem might all join a discussion on ways to solve the problem.
- Only a few knowledgeable individuals might give information when individuals seek to find out more about specific topics.



Introduction (2/2)

- In this study, we propose
 - a) a method of **social network extraction** from threaded messages,
 - b) “**Communication Gap**” as an indicator of understanding the state of communication,
 - c) to reveal the **characteristics of Communication Gap** by analyzing 3,000 social networks, and
 - d) an approach for **re-organizing communication structures** to realize innovative communication, based on the strategy for reducing Communication Gap.



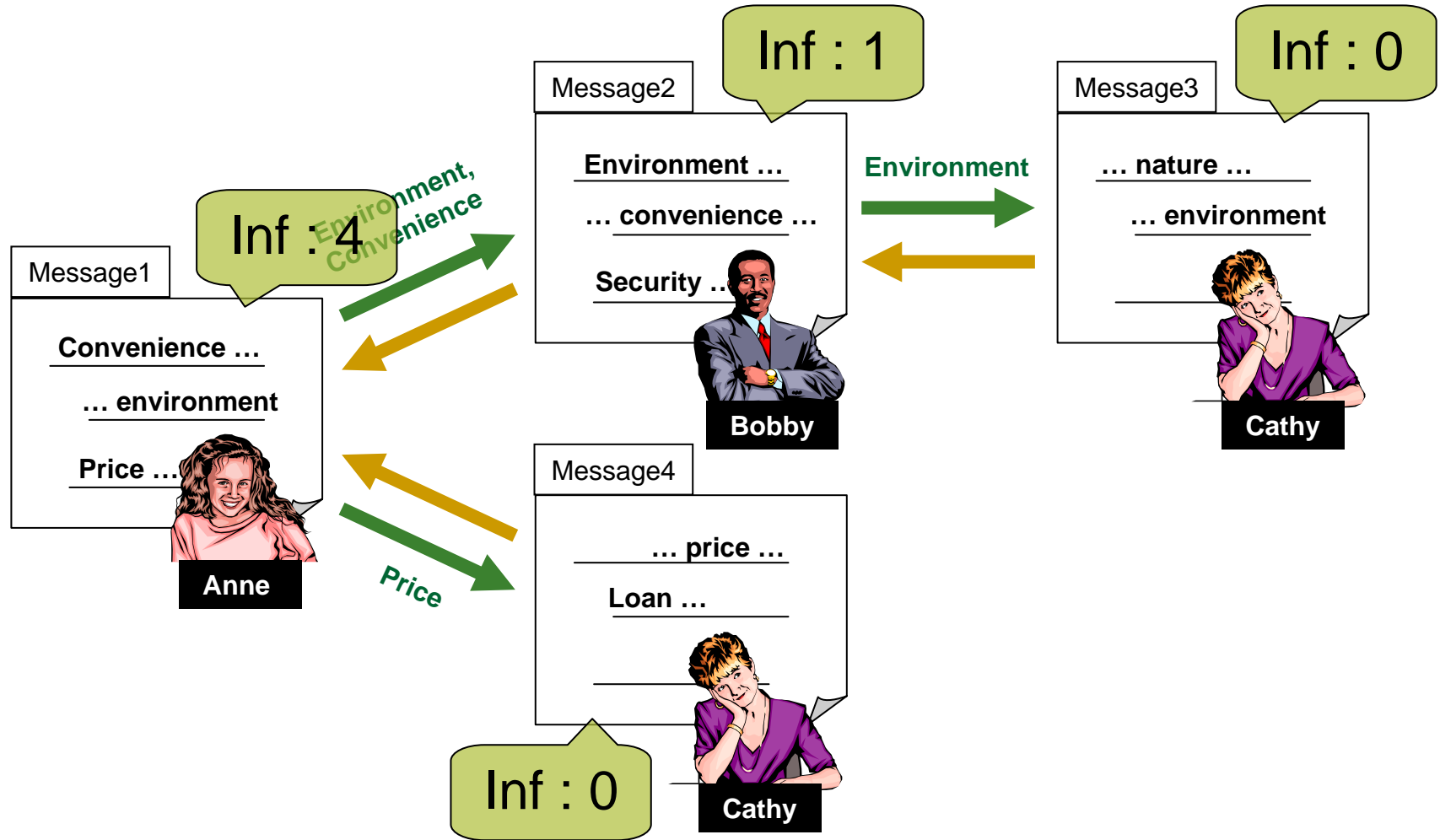
Previous Works

- Network analysis
 - Small world phenomena [Milgram 1967]
 - Social network analysis [Freeman 1973]
 - Information flow within an organization [Scott 1992]
 - Importance of informal networks [Krackhard et al.1993]
 - Collective dynamics of 'small-world' networks [Watts et al. 1998]
 - Dependency Structure Matrix Analysis [Tian-Li et al. 2004]
 - Communication analysis
 - Diffusion of Innovation Research [Rogers 1962]
 - GAs & Innovation Research [Goldberg 1983, 2002]
 - Chance Discovery [Ohsawa 2003]
 - Community of practice in E-mail logs [Tyler et al. 2003]
-

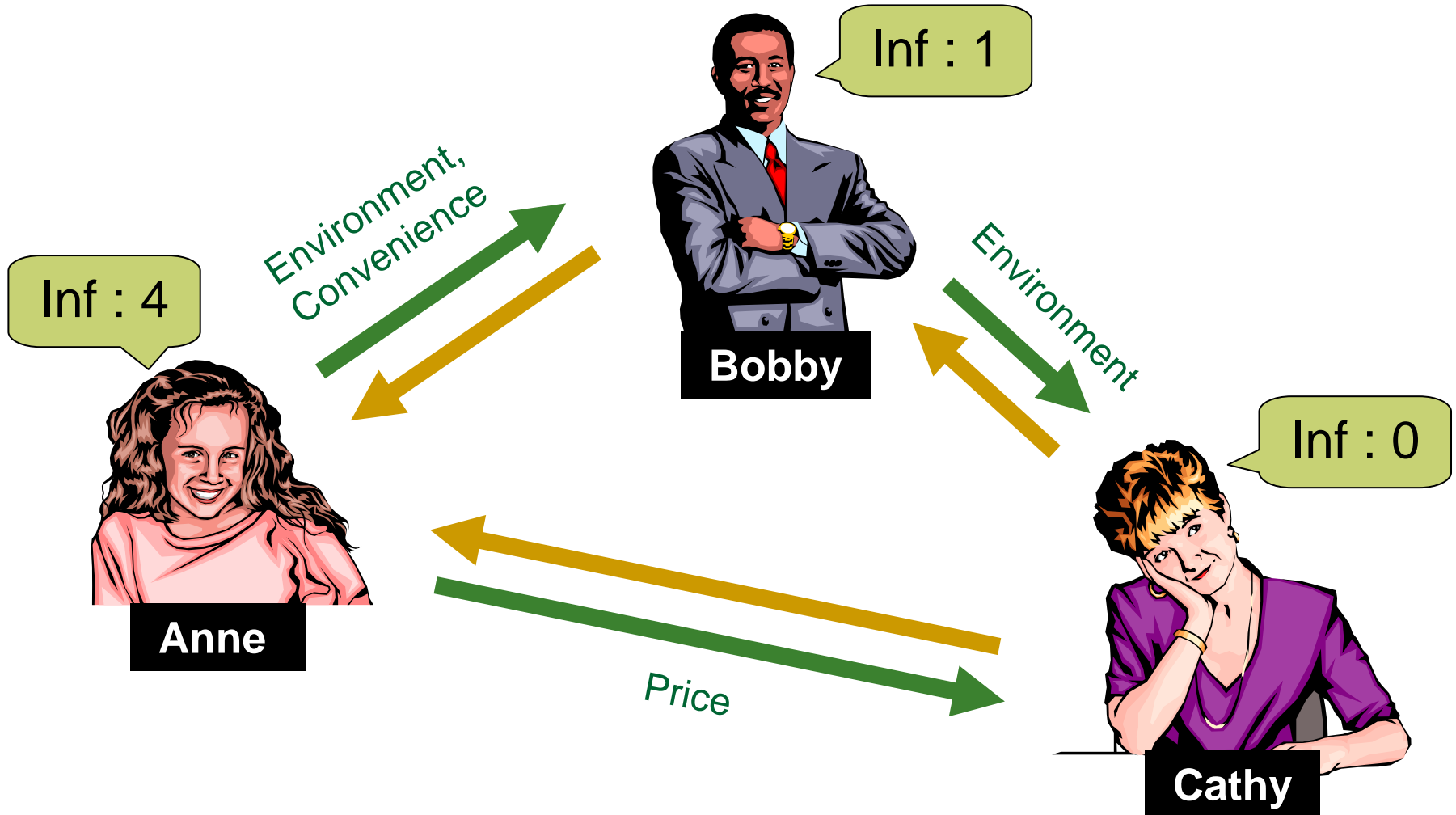
Extraction of Social Networks

- A social network here should be **directed** and **weighted** if we consider the flow of influence.
 - The relationship between individuals is not always symmetric because of their activities and social situations [Wallace 1999]
 - Directed and weight links are used when measuring Communication Gap (explained later).
 - We apply **IDM** (Influence Diffusion Model) to extract social networks from threaded messages [Matsumura 02]
 - IDM measures the influence of messages / individuals / terms / ties between individuals by focusing on propagating terms throughout message chains.
-

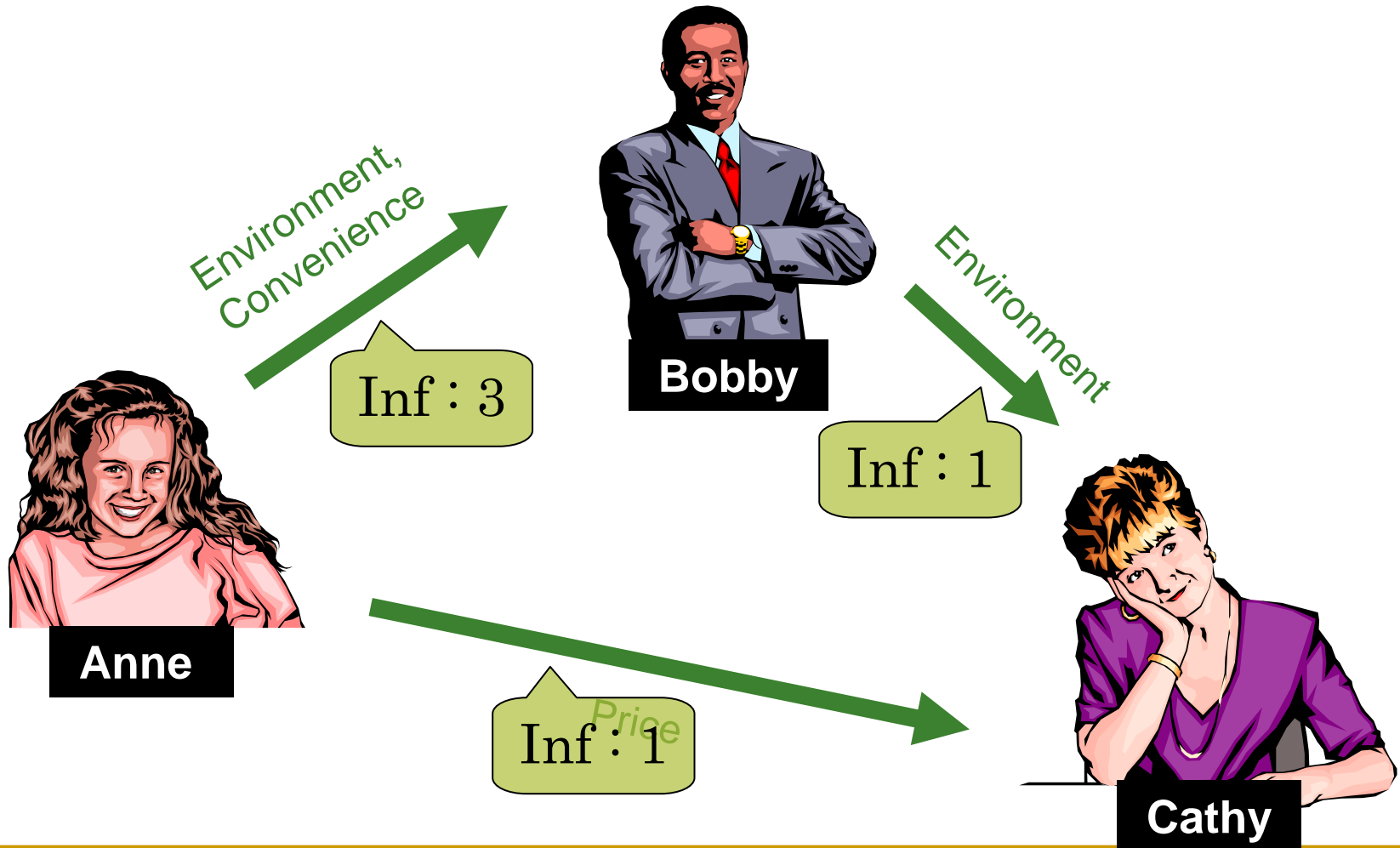
IDM (Influence Diffusion Model)



Influence of Individuals



Influence of Ties between Individuals

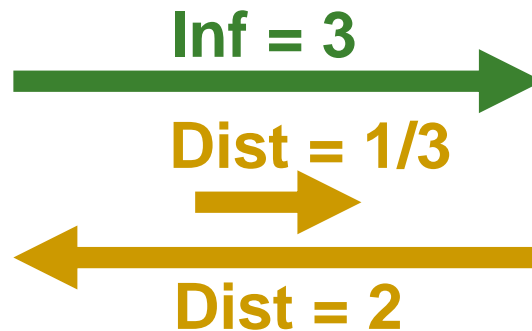


Distance from Influence

- As the influence shows contextual similarity between individuals, the distance could be defined as inversely proportionate to the influence, **Dist = 1 / Inf**.
- If there is no influence, the distance is defined as $n - 1$ (n is the number of individuals) as the case of the weakest relationships.

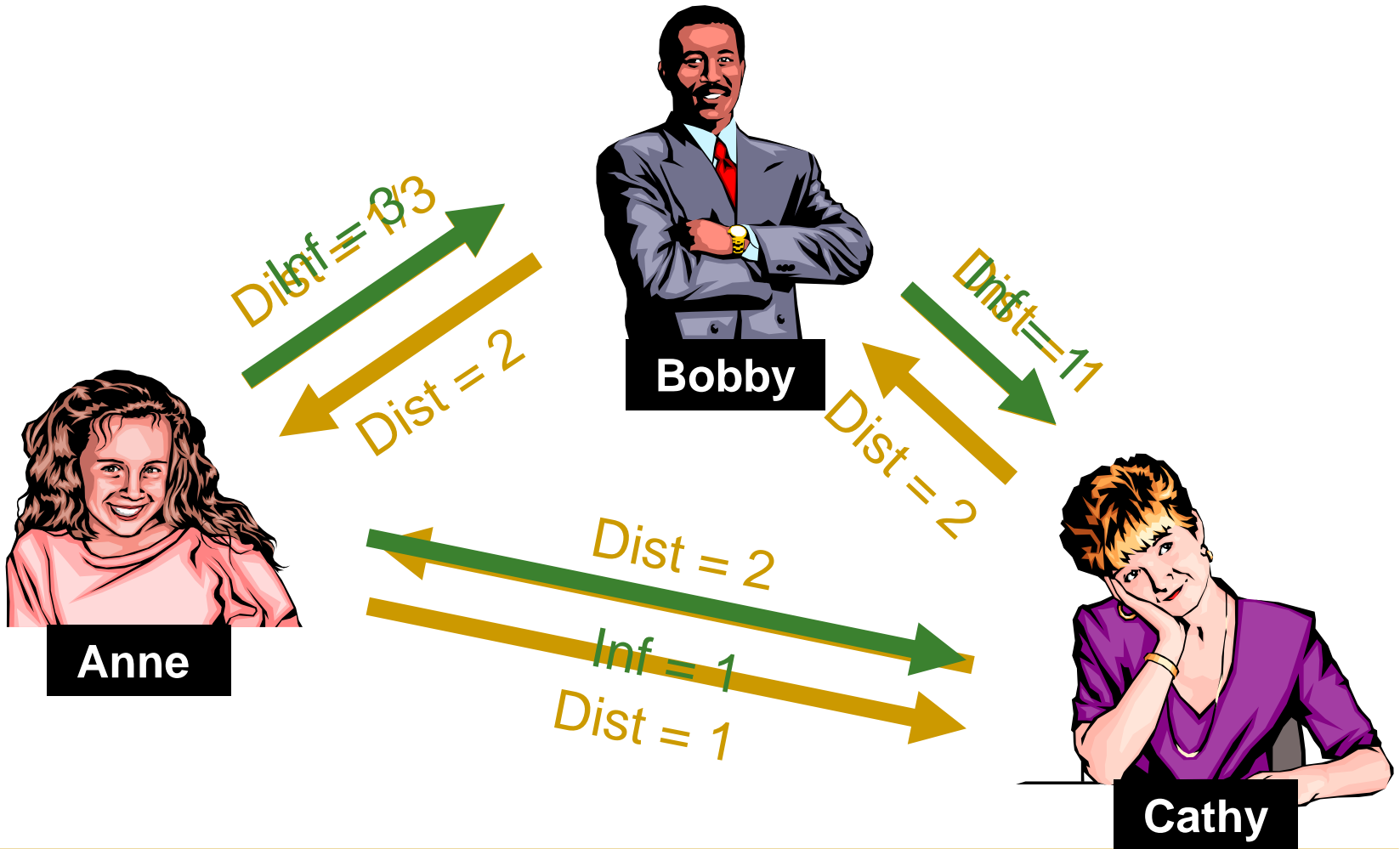


Anne



Bobby

Directed and Weighted Social Network



Communication Gap

- We measure **Communication Gap** from a social network to understand how much individuals are isolated in communication.
- Communication Gap is measured by
 - accumulating the differences of the shortest paths between individuals,
 - accumulating the longer/shorter distances of the shortest paths between individuals as a bottleneck hindering communication,
 - accumulating the differences in closeness centralities of individuals,
 - accumulating the closeness centralities of individuals.



Communication Gap Indices

- C = $\alpha_1 \sum_{p_x} c_{p_x}^{in}$
- C_{diff} = $\alpha_2 \sum_{p_x} |c_{p_x}^{in} - c_{p_x}^{out}|$
- G_{diff} = $\alpha_3 \sum_{p_x} \sum_{p_y} |d_{p_x, p_y} - d_{p_y, p_x}|$
- G_{max} = $\alpha_4 \sum_{p_x} \sum_{p_y} \max(d_{p_x, p_y}, d_{p_y, p_x})$
- G_{min} = $\alpha_5 \sum_{p_x} \sum_{p_y} \min(d_{p_x, p_y}, d_{p_y, p_x})$

$c_{p_x}^{in}$: Closeness centrality of p_x (inward)
 $c_{p_x}^{out}$: Closeness centrality of p_x (outward)
 d_{p_x, p_y} : The distance from p_x to p_y
 α_n : Normalization coefficient
 p_x, p_y : An individual in a social network

Yahoo! JAPAN BBS

We downloaded 3,000 BBS and extracted 3,000 social networks!

The image displays three overlapping screenshots of the Yahoo! JAPAN website and BBS interface, likely from the year 2000, as indicated by the dates in the screenshots.

The top-left screenshot shows the main Yahoo! JAPAN homepage with the search bar and navigation links.

The middle-left screenshot shows the BBS category selection page, listing various topics such as Entertainment, Sports, Business, and Life & Culture.

The right screenshot shows a BBS message thread titled "瓢箪山の方いらっしゃいますか?". The thread contains several replies with timestamps and user names.

タイトル	投稿者	日時
=> 瓢箪山の方いらっしゃいますか?	hdftu458	2000/6/29 4:10
└はい!	yassanman	2000/7/2 0:59
└池島高校	yamasan1013	2003/2/21 0:47
└私は山本です。	mugu88	2000/7/8 23:56
└いらっしゃ〜!	yassanman	2000/7/9 1:39
└去年まで、住んでました	melt-downmelt-down	2000/7/10 4:37
└こんちわ〜	missusite	2000/7/15 18:58
└長崎屋は...	yassanman	2000/7/15 21:07
└神田町に住む友へ	dvm52291	2000/7/15 21:52
└お久しぶりです("o")	mugu88	2000/7/16 20:32
└[さらにスレッドを見る]		
└こんばんはー	mugu88	2000/7/16 20:47
└懐かしい...	tosiori	2000/7/18 14:25
└はじめまして	mitsuo0083	2000/7/18 16:09
└今日は稲荷まつり!	yassanman	2000/7/18 21:53
└こんばんは!!	nyanmage-mura	2000/7/19 1:21
└おは!	missusite	2000/7/19 7:54
└こんち	missusite	2000/7/19 15:23
└ういっす!	yassanman	2000/7/20 0:59

Communication Gaps in Y!Japan BBS

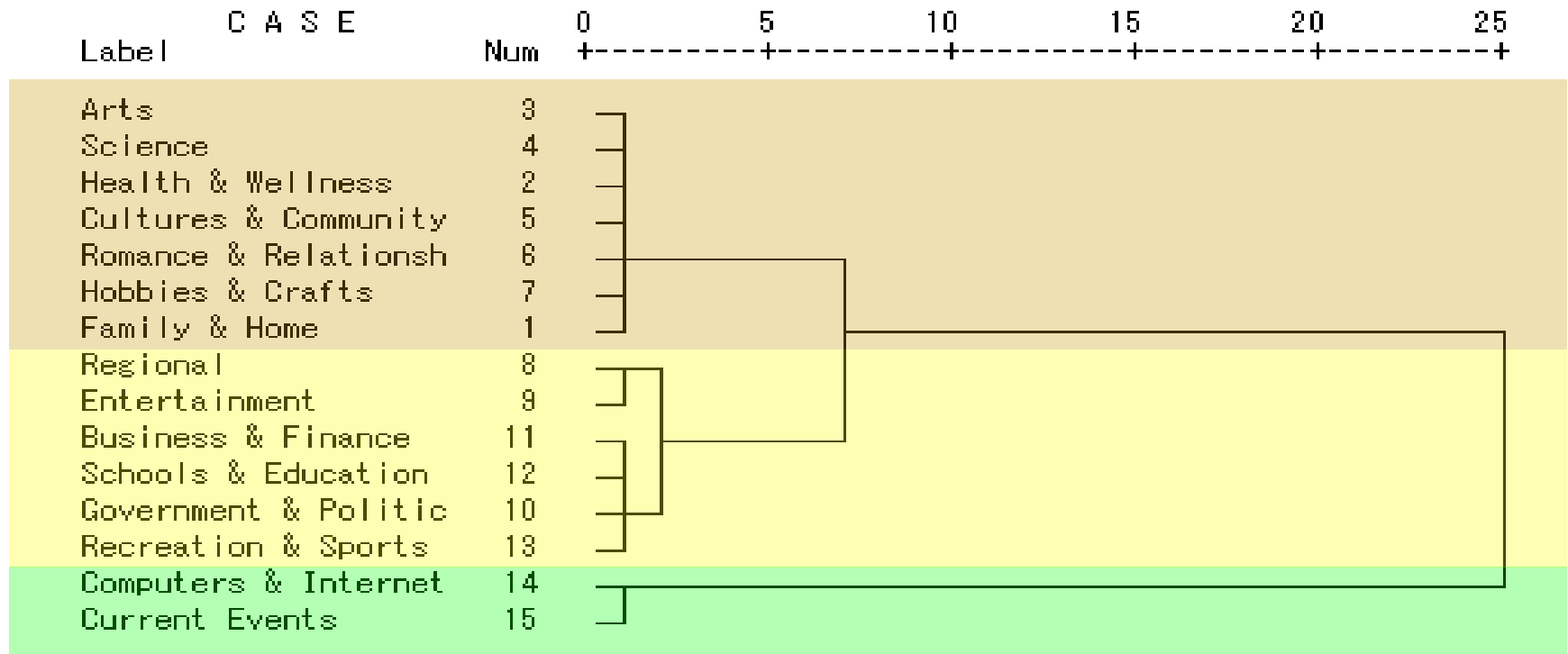
Yahoo!Japan Categories	C_dist	C_diff	G_diff	G_max	G_min
Family & Home	0.032	0.028	0.032	0.048	0.017
Health & Wellness	0.065	0.065	0.073	0.100	0.029
Arts	0.068	0.070	0.081	0.108	0.029
Science	0.072	0.068	0.078	0.110	0.034
Cultures & Community	0.085	0.061	0.071	0.120	0.051
Romance & Relationships	0.081	0.079	0.089	0.125	0.038
Hobbies & Crafts	0.095	0.092	0.106	0.146	0.043
Regional	0.161	0.120	0.142	0.230	0.093
Entertainment	0.151	0.129	0.157	0.228	0.075
Government & Politics	0.217	0.167	0.197	0.313	0.120
Business & Finance	0.241	0.160	0.184	0.331	0.150
Schools & Education	0.253	0.161	0.195	0.349	0.158
Recreation & Sports	0.239	0.208	0.253	0.362	0.116
Computers & Internet	0.447	0.221	0.272	0.579	0.315
Current Events	0.455	0.220	0.271	0.588	0.322

Three Types of Communication

***** H I E R A R C H I C A L C L U S T E R A N A L Y S I S *****

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine



Three Types of Communication

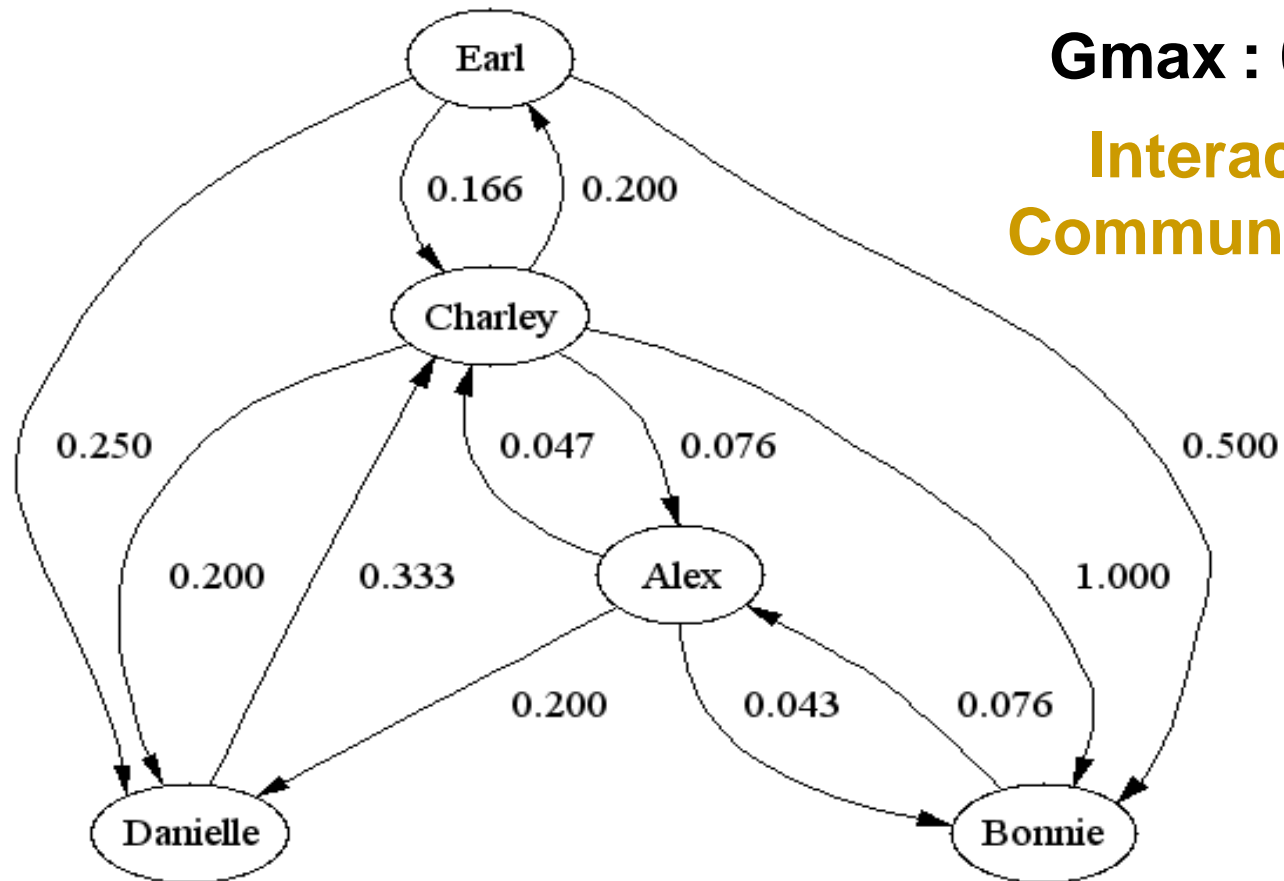
Yahoo!Japan Categories	C_dist	C_diff	G_diff	G_max	G_min
Family & Home	0.032	0.028	0.032	0.048	0.017
Health & Wellness	0.065	0.065	0.073	0.100	0.029
Arts	0.068	0.070	0.081	0.108	0.029
Science	0.072	0.068	0.078	0.110	0.034
Cultures & Community	0.085	0.061	0.071	0.120	0.051
Romance & Relationships	0.081	0.079	0.089	0.125	0.038
Hobbies & Crafts	0.095	0.092	0.106	0.146	0.043
Regional	0.161	0.120	0.142	0.230	0.093
Entertainment	0.151	0.129	0.157	0.228	0.075
Government & Politics	0.217	0.167	0.197	0.313	0.120
Business & Finance	0.241	0.169	0.184	0.331	0.150
Schools & Education	0.253	0.161	0.195	0.349	0.158
Recreation & Sports	0.239	0.208	0.253	0.362	0.116
Computers & Internet	0.447	0.221	0.272	0.579	0.315
Current Events	0.455	0.220	0.271	0.588	0.322

Interactive Communication

Distributed Expertise Communication

Soapbox Communication

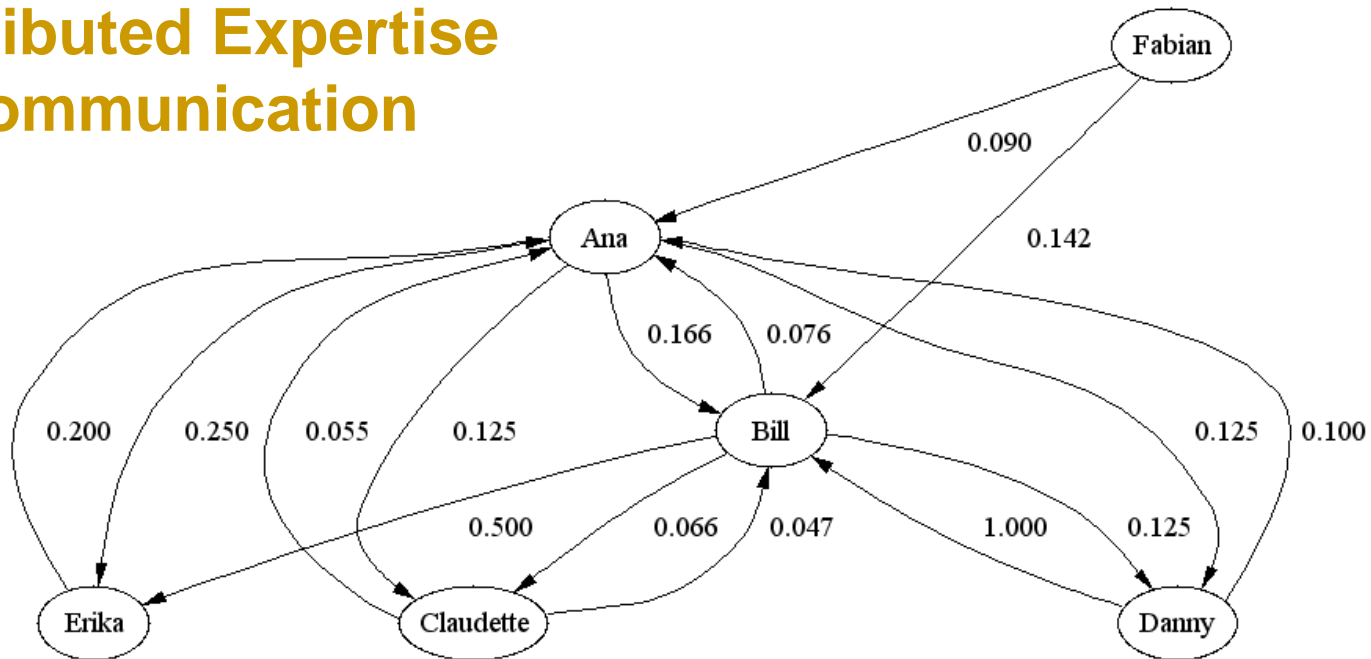
Case 1: "Cell Phone and Women"



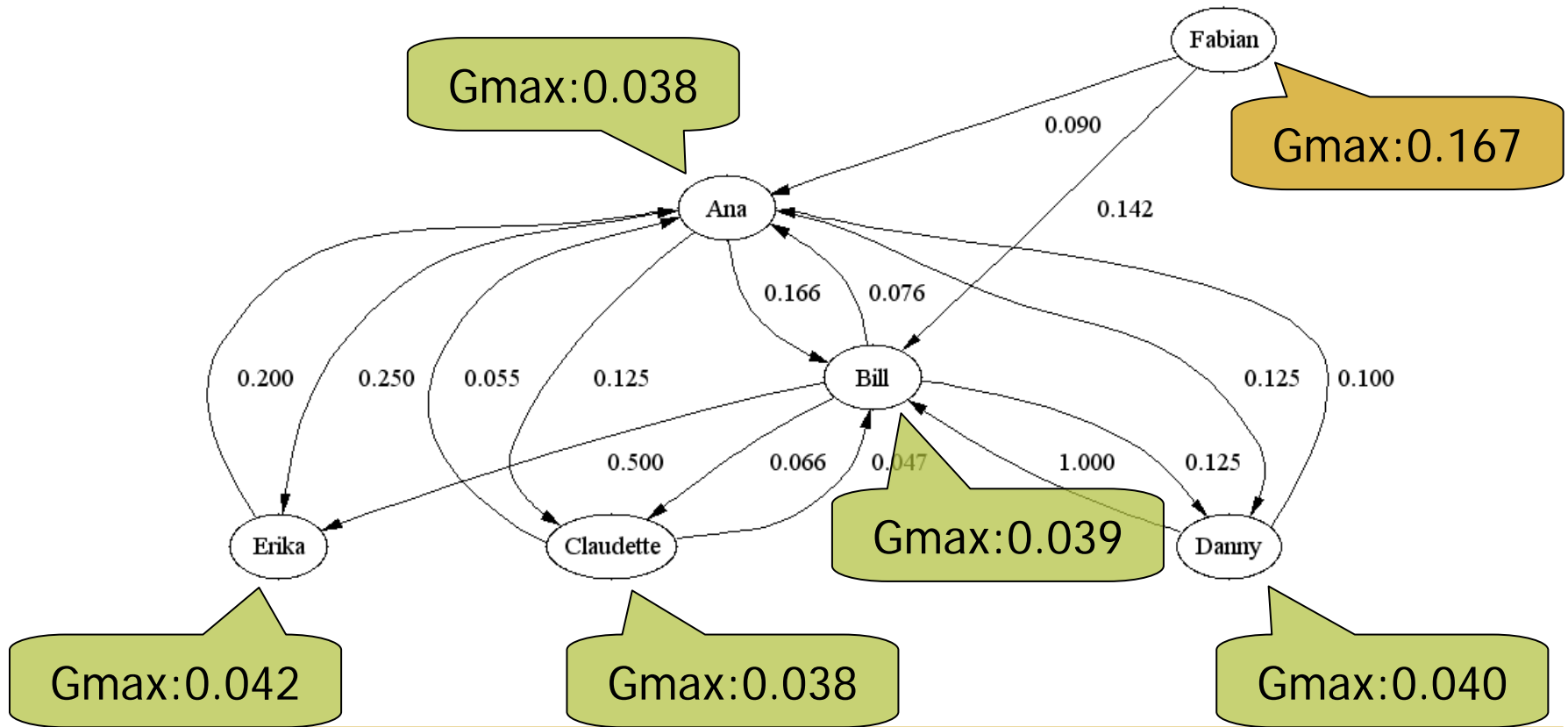
Case 2: "Purchase of a House"

Gmax : 0.364

Distributed Expertise Communication

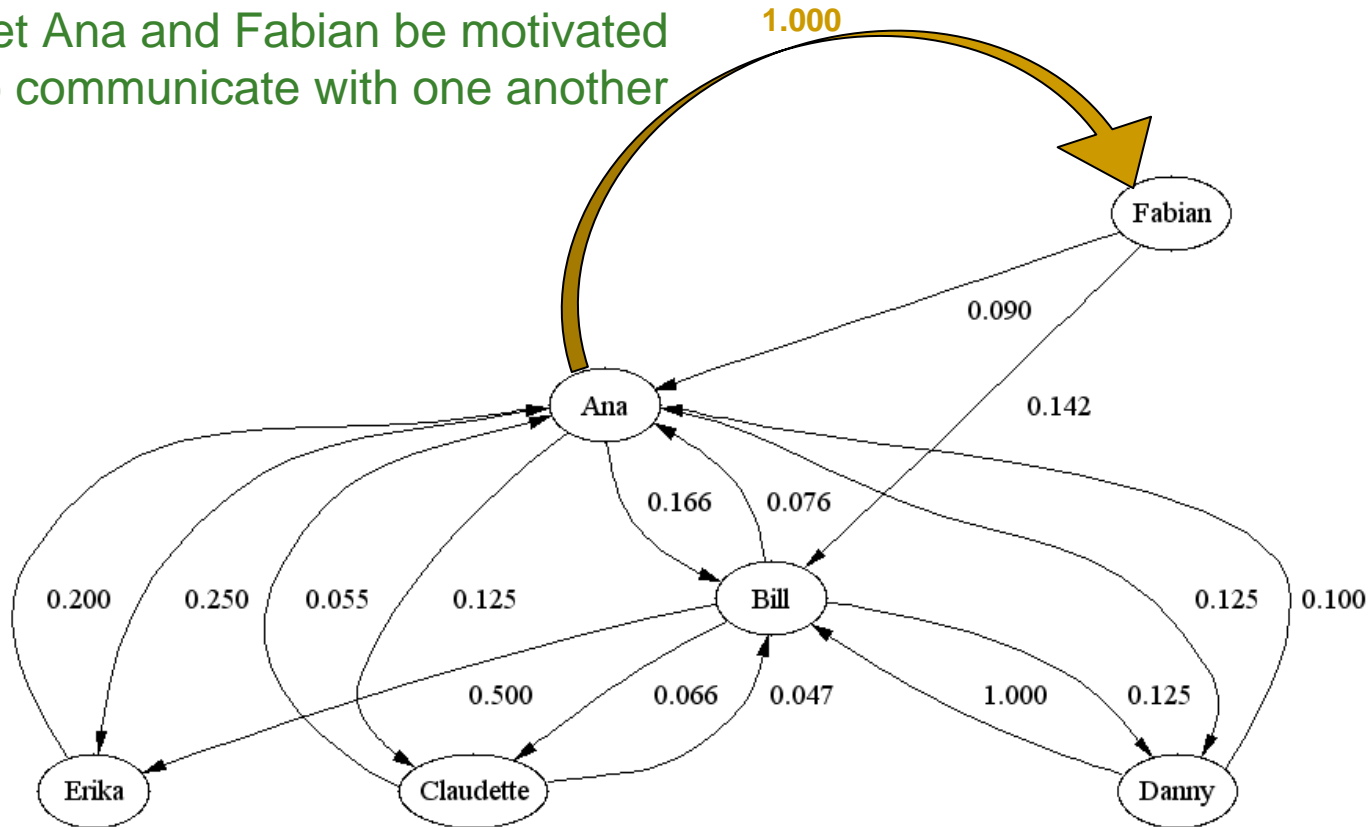


Communication Gaps of Individuals



Remedy 1

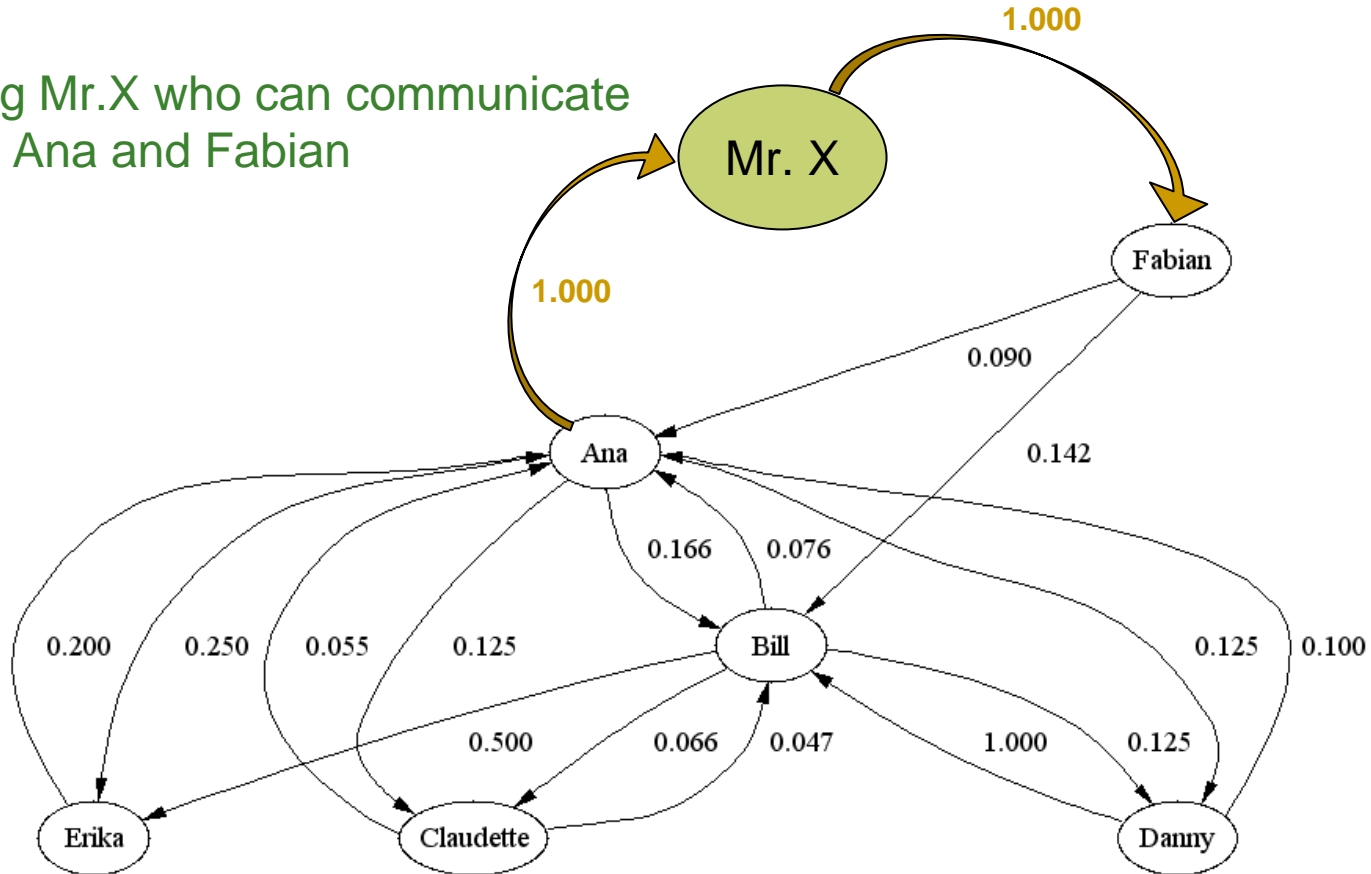
Let Ana and Fabian be motivated to communicate with one another



Gmax = 0.364 (Distributed Expertise Communication)
→ 0.103 (Interactive Communication)

Remedy 2

Bring Mr.X who can communicate with Ana and Fabian



Gmax = 0.364 (Distributed Expertise Communication)
→ 0.157 (Interactive Communication)

Conclusion and Future Direction

- We have proposed a method to re-organize communication structures for innovative communication based on Communication Gap in a social network.
- IDM has already implemented on DISCUS (Distributed Innovation and Scalable Collaboration in Uncertain Settings) platform developed at IlliGAL and ALG/NCSA.
- As a future direction, we will apply Communication Gap methodology on DISCUS platform to re-organize the communication structures and accelerate the emergence of human innovation.

