A Computational Framework for Social Capital in Online Communities

Matthew Smith

http://m.smithworx.com

Department of Computer Science

http://dml.cs.byu.edu

Brigham Young University

DML

Data Mining Lab

PhD Dissertation Defense, April 2011

Introduction

- Online communities increasingly important
- Prevalent shift in how people discover information
- Large social networks are dynamic and complex
- Rich social network data is becoming available
- Social capital within these networks is poorly understood
- **Computational Social Science**

Social Capital Framework

Social Capital

- **The value of social networks.** More specifically,
  - ***Bonding*** similar people together (Putnam)
  - ***Bridging*** diverse people together (Putnam)
  - *Access* to and *use of **resources*** embedded in social networks (Lin)

Social Capital

- **Relationships**
  Burt, Portes
- **Affinities**
  Putnam, Coleman
- **Social Resources**
  Lin, Bourdieu

Social Capital

Relationships

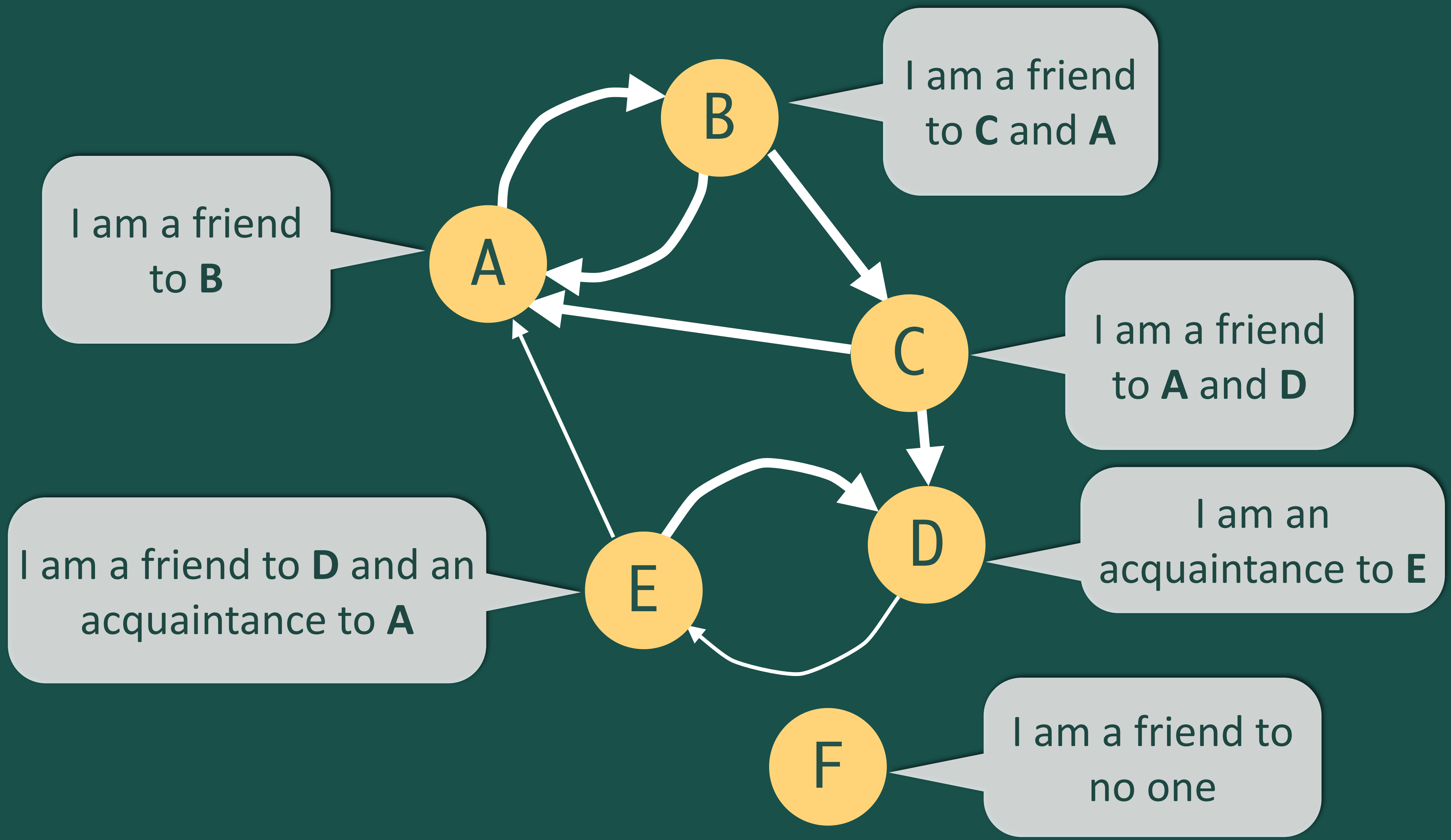
- **Explicit Link**

  - Direct knowledge, interaction, or communication
  - Ex. behavioral interaction, evaluation of one by another (e.g., friends, web links), formal (e.g., co-worker), biological (e.g., sibling)

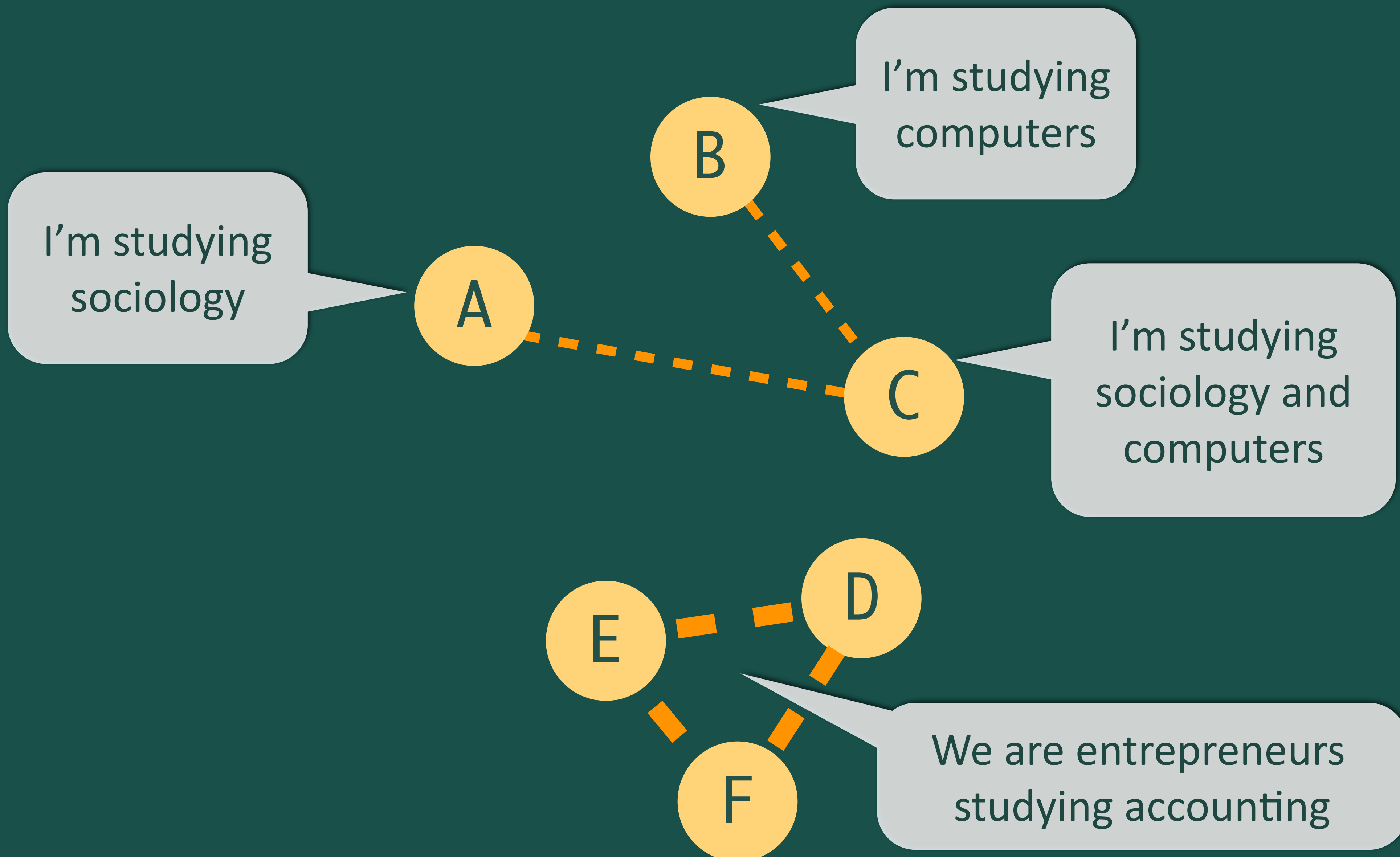
- **Implicit Link**

  - Inherent similarities or affinities
  - Ex. attributes, hobbies, interests, background, etc.

Explicit Social Network (ESN)

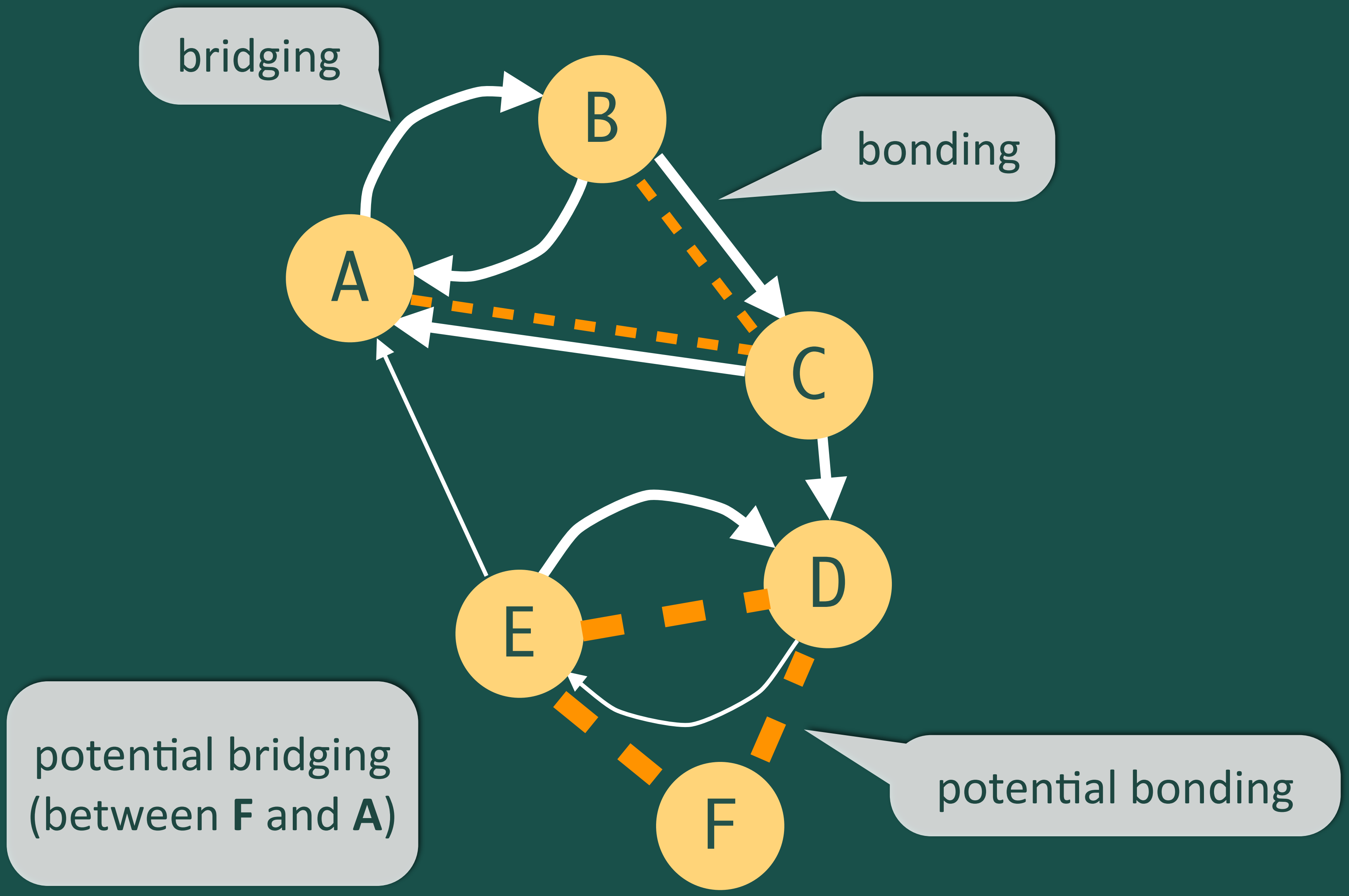


Implicit Affinity Network (IAN)



Smith, M., Giraud-Carrier, C. and Judkins, B. (2007). Implicit Affinity Networks. In Proceedings of the Seventeenth Annual Workshop on Information Technologies and Systems, 1-6.

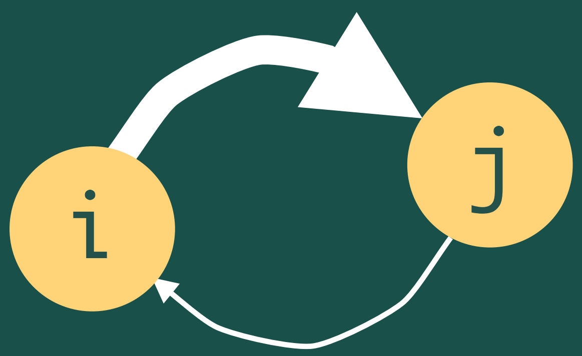
Hybrid Network



Smith, M., Giraud-Carrier, C., and Purser, N. (2009). Implicit Affinity Networks and Social Capital. Information Technology and Management, 10(2-3):123-134, September.

Relationships

- Directed



- Strength
  - *Initialization*: Behavioral interaction, evaluation of one person by another, formal relationships, biological relationships; base on affinities
  - *Dynamic*: interactions, time decay

Affinities

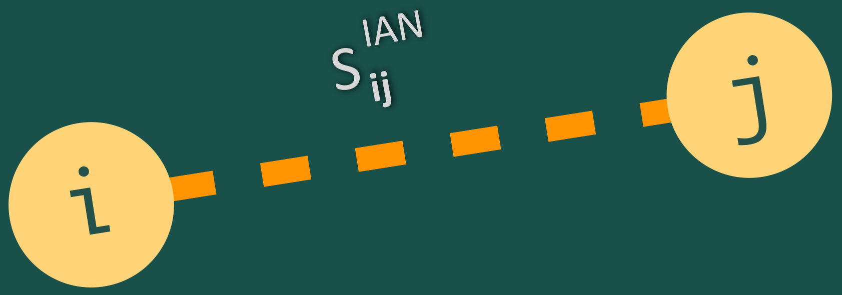
- The set of attributes exposed by individual ***j***

$$A_j$$

- The affinity strength between individuals ***i*** and ***j***

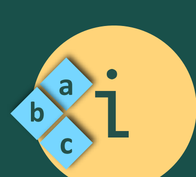
$$s_{ij}^{IAN} = \frac{A_i \cap A_j}{A_i \cup A_j}$$

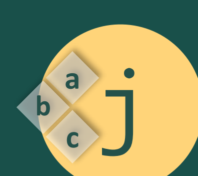
Jaccard Index
(set similarity)



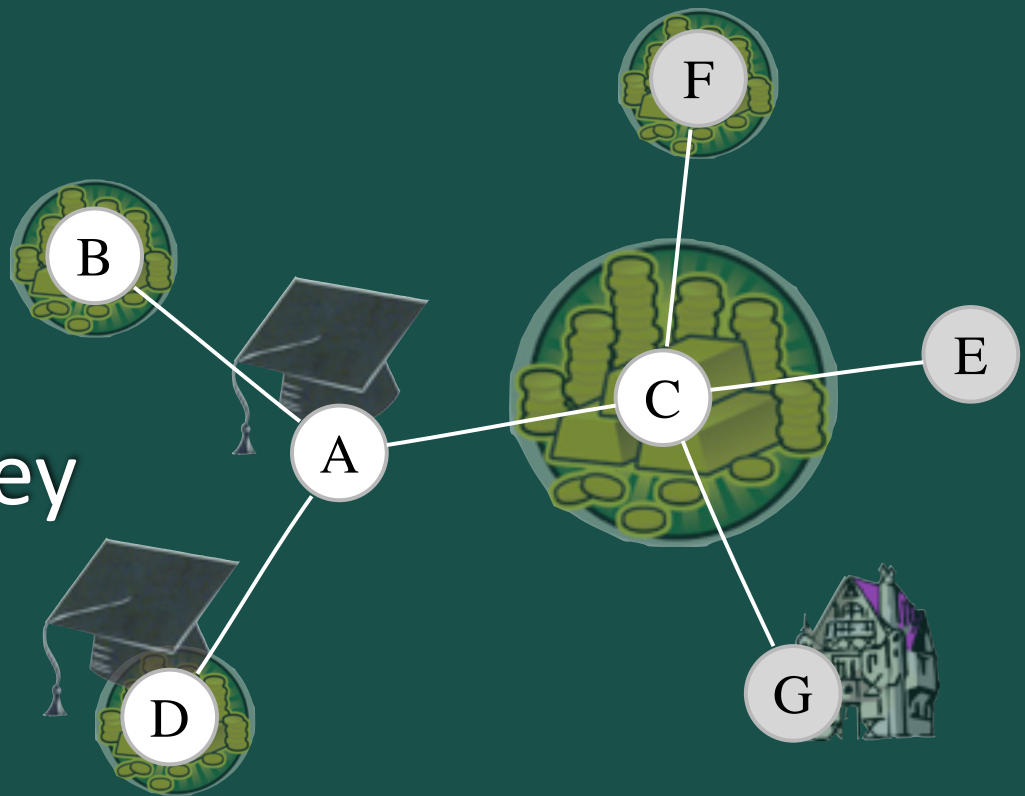
Resources

- ***Social* resource:** a specific asset, material or symbolic, available through social connections within a network
- *Characteristics*: (e.g., exhaustible, returnable, quantifiable, durable)
- *Value*: assigned by the individual (often dependent on characteristics)
- *Possessed* or *Sought*





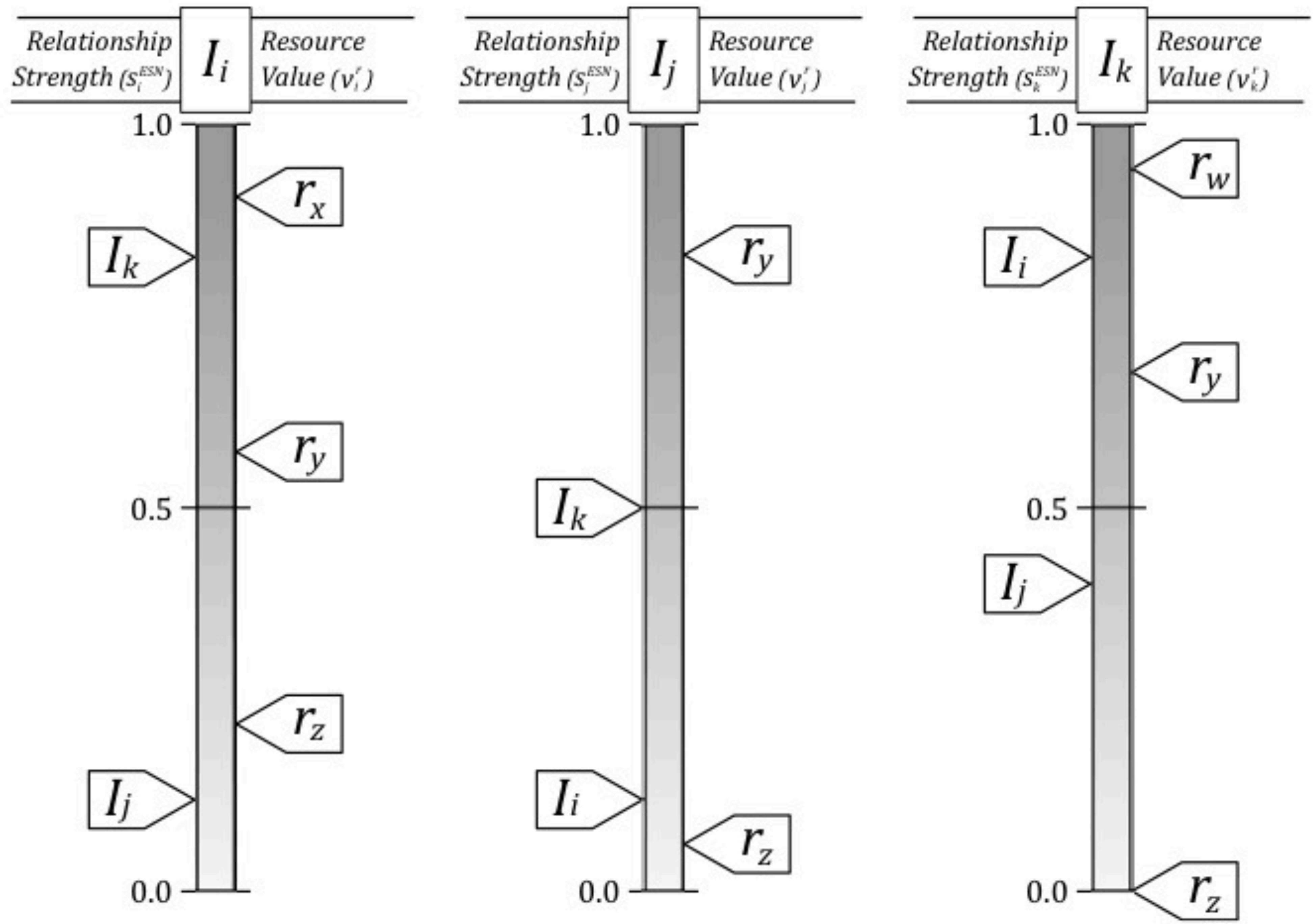
Social Resource Examples



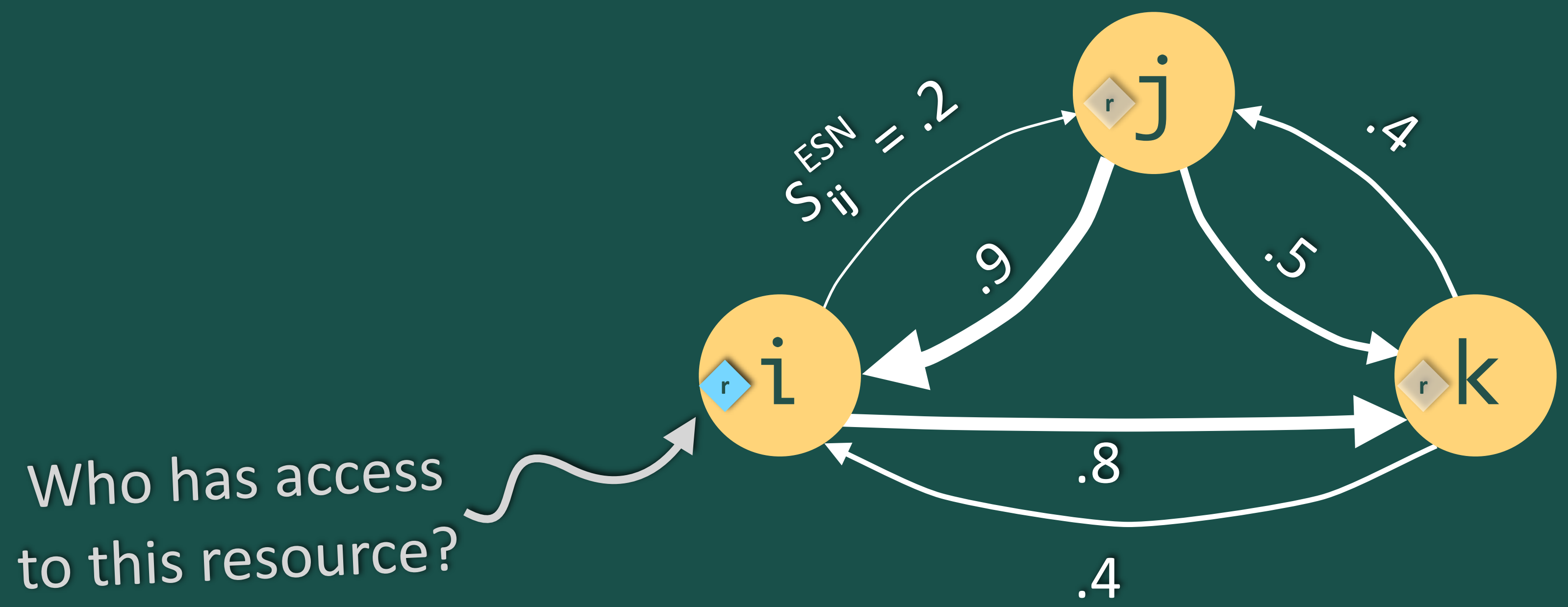
- Material Goods
  - land, houses, car, and money
- Symbolic Goods
  - education, memberships in clubs, honorific degrees, nobility or organizational titles, family name, reputation, or fame

From: Lin 2001, Social Capital: A Theory of Social Structure and Action

Access to Resources



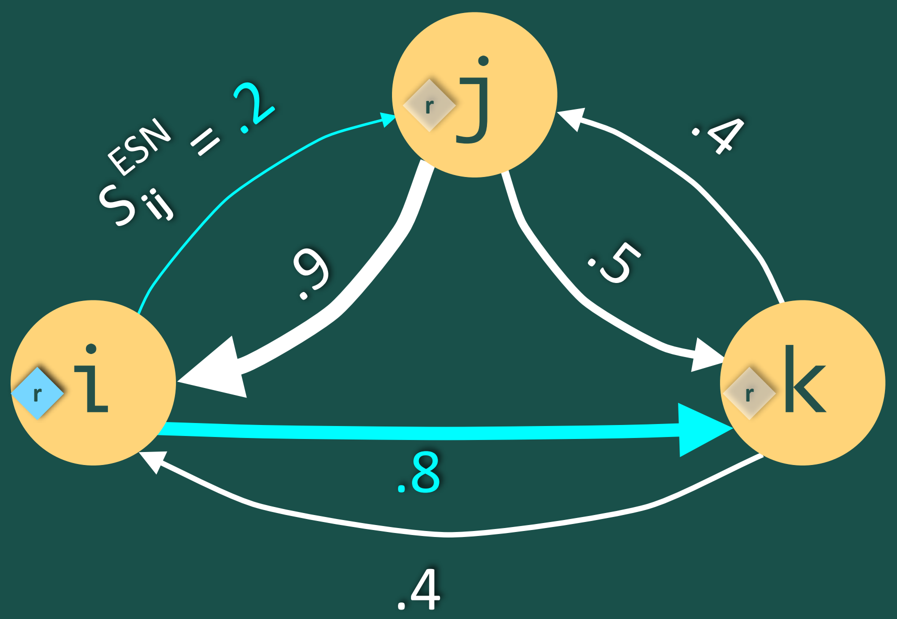
Access to Resources



$v_i^r = .4$

$$access_j(i,r) = \begin{cases} True & \text{if } s_{ij}^{ESN} \geq v_i^r \\ False & \text{otherwise} \end{cases}$$

Access to Resources



Only individual ***k*** has access

v_i^r = .4

$$access_j(i,r) = \begin{cases} True & \text{if } s_{ij}^{ESN} \geq v_i^r \\ False & \text{otherwise} \end{cases}$$

Social Capital Computation

- Resources available to individual ***j*** from individual ***i***

$$\#res_j(i) = |\ \{r \in R_i^p : access_j(i,r) = True \wedge C_j(r) = Yes\}\ |$$

Context Vector

- Individual Social capital

$$sc(j) = \sum_i \#res_j(i)$$

Bonding and Bridging Social Capital

- Bonding and Bridging

$$b(j) = \sum_i \#res_j(i) s_{ij}^{IAN}$$

$$br(j) = \sum_i \#res_j(i)(1 - s_{ij}^{IAN})$$

- Individual

$$sc(j) = b(j) + br(j)$$

- Network

$$sc = \sum_j sc(j)$$

Bonding and Bridging Social Capital

- Bonding and Bridging

$$b(j) = \sum_i s_{ij}^{ESN} s_{ij}^{IAN}$$

$$br(j) = \sum_i s_{ij}^{ESN} (1 - s_{ij}^{IAN})$$

Note: used as surrogate when resources are not available

- Individual

$$sc(j) = b(j) + br(j)$$

"It is a very interesting implementation of the social capital ideas, especially *bridging and bonding*. I have long thought that the social capital research field would be immensely aided if some of it could be formalized in terms of network theory, and this is **the best job of that that I've seen so far**."

- Robert Putnam *(personal correspondence)*

Interactions

- **Interaction:** a purposeful exchange between individuals
  - How is it perceived? (+, -, or neutral)
  - Is a resource involved?

- Who to interact with?

Interactions

- Generalized form

$$i = sel_j(R_i^p, R_j^s, s_{ij}^{ESN}, s_{ij}^{IAN})$$

Resources possessed by **i**

Resources sought by **j**

- Instantiated examples (determined by ***j***'s goals)

$$i = argmin_k(min_{r \in R_j^s \cap R_k^p}(v_k^r - s_{kj}^{ESN}))$$

Get sought after resources as quickly as possible

$$i = argmin_k(v_k^{r_h} - s_{kj}^{ESN})$$

Seek resources in a priority order

$$i = argmax_k(| R_k^p | + s_{kj}^{IAN})$$

Seek resources from who has many and is most similar to you

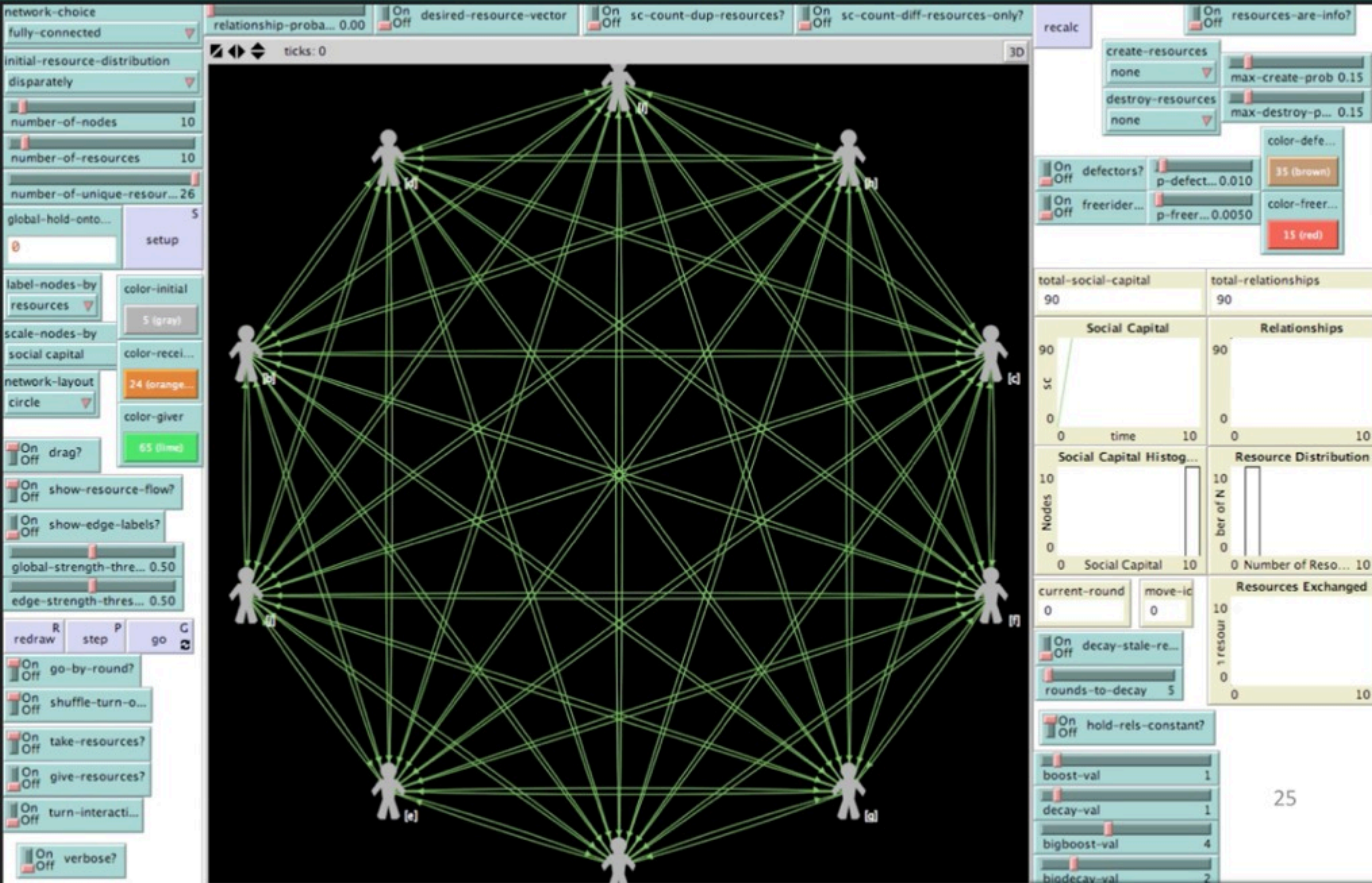
Dynamic Interaction Function

$$\Delta s_{ji}^{ESN}(event)$$

Event	Δs_{ji}^{ESN}	Δs_{ij}^{ESN}
j gives resource r to i	$-$	$+$
j receives resource r from i	$+$	$-$
j engages in a (resourceless) interaction with i	$-$	$+$
j receives a (resourceless) interaction from i	$+$	$-$

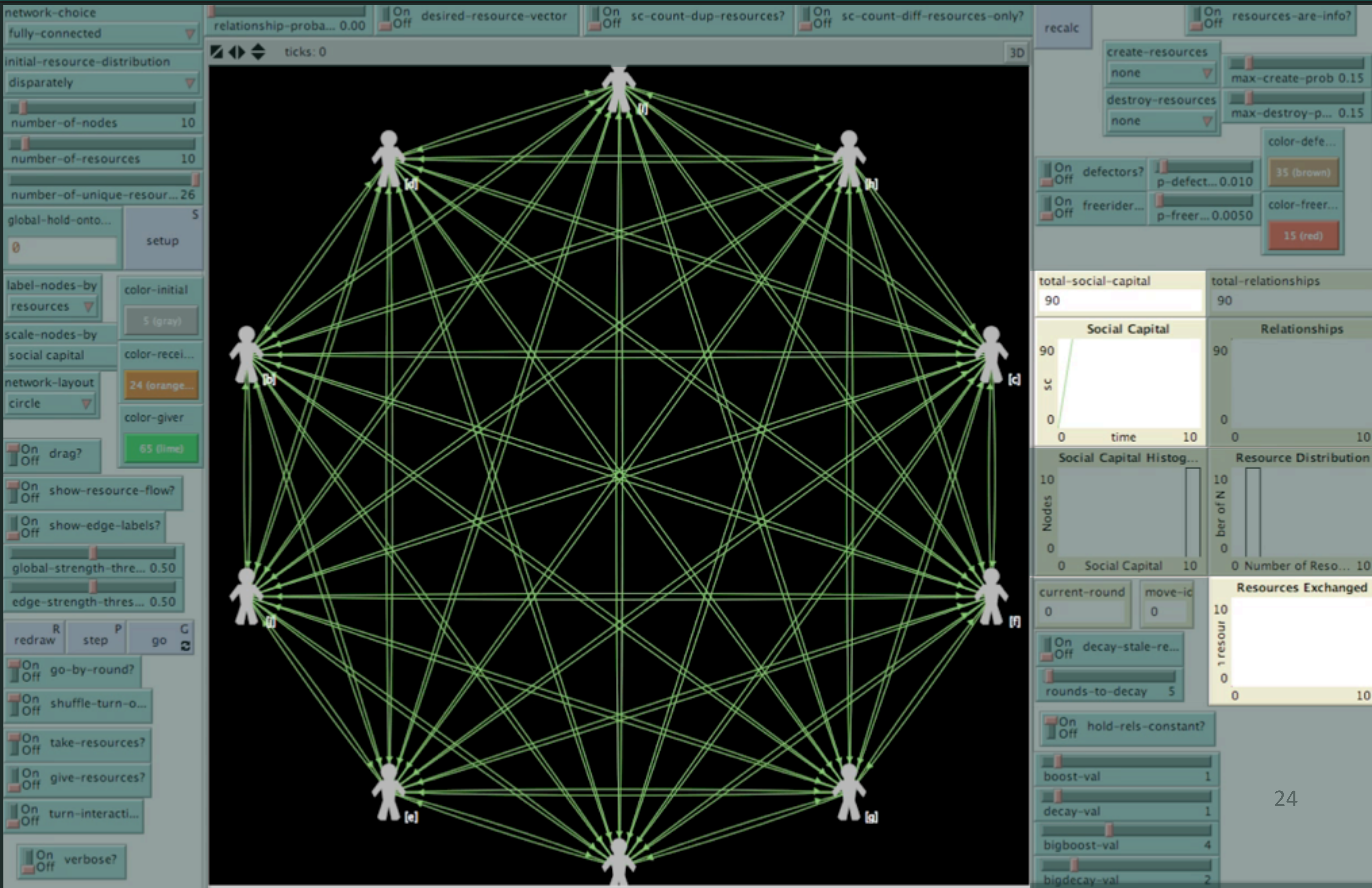
Implementation

NetLogo



Simulation

Community & defectors



Publications (Framework)

- Implicit Affinity Networks
  (published in *Proceedings of the 17th Annual Workshop on Information Technologies and Systems*)

- Implicit Affinity Networks and Social Capital
  (published in *Information Technology and Management - Journal*)

- Measuring and Reasoning about Social Capital
  (submitted to *Social Networks*, 2011)

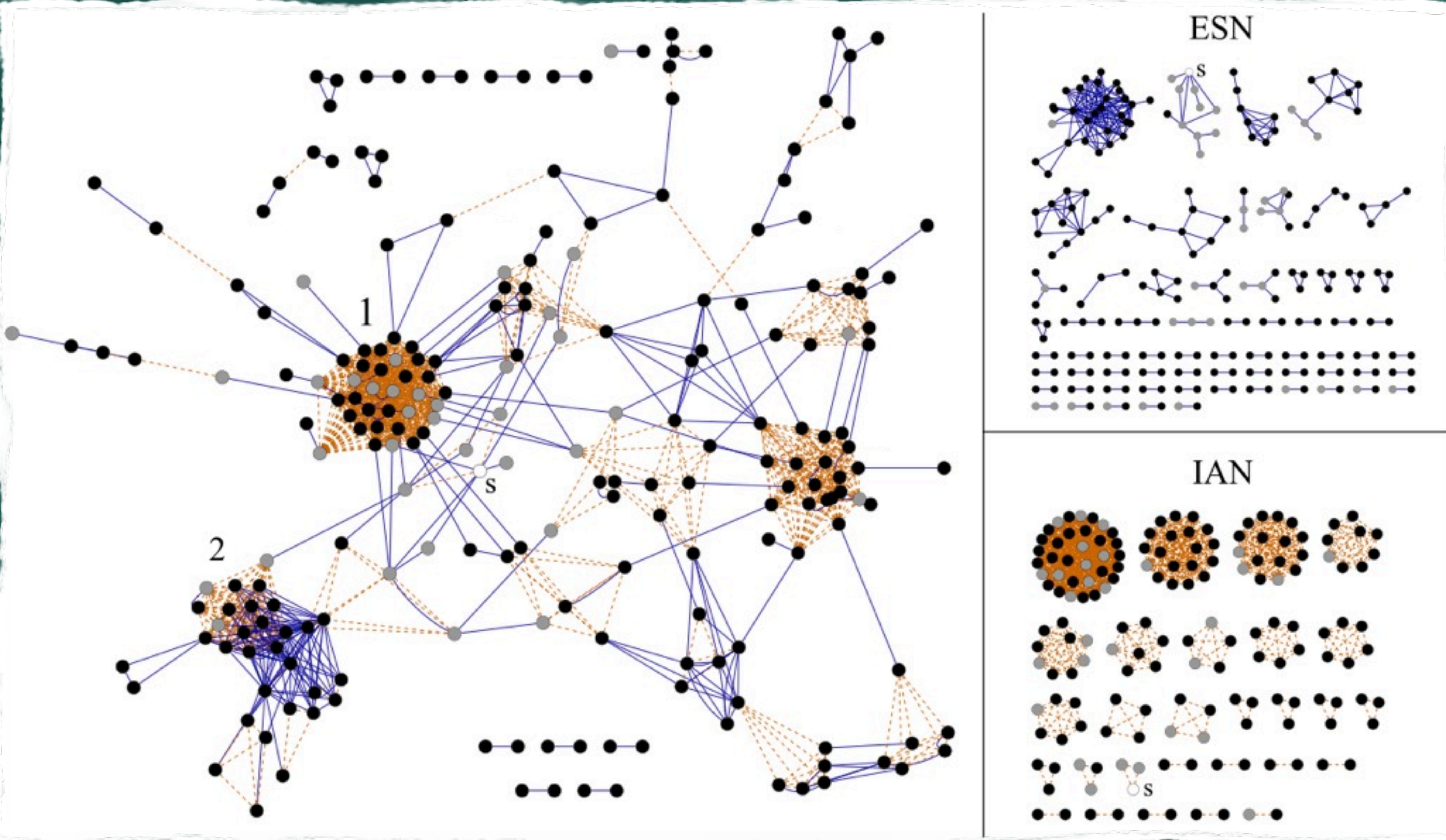
Case Studies

Blogosphere

Blogosphere Experiment

- Focus
  - Blogger communities centered around topics
  - Started with Scoble's blog
- Data
  - 13 million blog entires, 38K+ blogs
  - July 2006 - July 2007 (1 year)

Hybrid Network - Blogosphere



Twitter

Twitter Experiment

- Leverages the Framework to empirically test:
  - “Bonding is more likely to occur than bridging” - Lin 01
  - “Closure is the most obvious force” - Burt 05
  - Homophily principle: “Birds of a feather flock together”
    Smith-Lovin 87 “Similarity begets friendship”

- *Following users with whom the most affinities are shared (i.e., attempting to bond) produces more follow-backs (i.e., bonding) than other following strategies.*

Results

rank_{fb}	strategy	following	↓ follow-backs	followers	rejects	churn
1	*bonding* (A)	500	158 (32%)	202 (40%)	12	127
2	*random* (D)	500	118 (24%)	154 (31%)	20	103
3	*median affinities* (C)	500	99 (20%)	123 (25%)	25	93
4	*bridging* (B)	500	99 (20%)	120 (24%)	25	91
5	*min. following/ers diff.* (E)	500	87 (17%)	99 (20%)	50	55
6	*max. following/ers diff.* (F)	500	84 (17%)	172 (34%)	12	324
7	*median num. followers* (G)	500	63 (13%)	86 (17%)	31	51
8	*min. num. followers* (H)	500	33 (07%)	42 (08%)	79	29
9	*follow nobody* (I)	0	0 (—%)	3 (—%)	0	24

strategy	significantly different
(A) *bonding*	B, C, E, F, G, H
(B) *bridging*	A, H
(C) *median affinities*	A, G, H
(D) *random*	G, H
(E) *min. following/ers diff.*	A, H
(F) *max. following/ers diff.*	A, H
(G) *median num. followers*	A, D, C
(H) *min. num. followers*	A, B, C, D, E, F

TABLE III
FOLLOWBACKS-TO-FOLLOWING: PAIRWISE PROPORTION TEST RESULTS. ($\alpha = 0.01$, BONFERRONI CORRECTED p-VALUES)

Follow-backs Retained

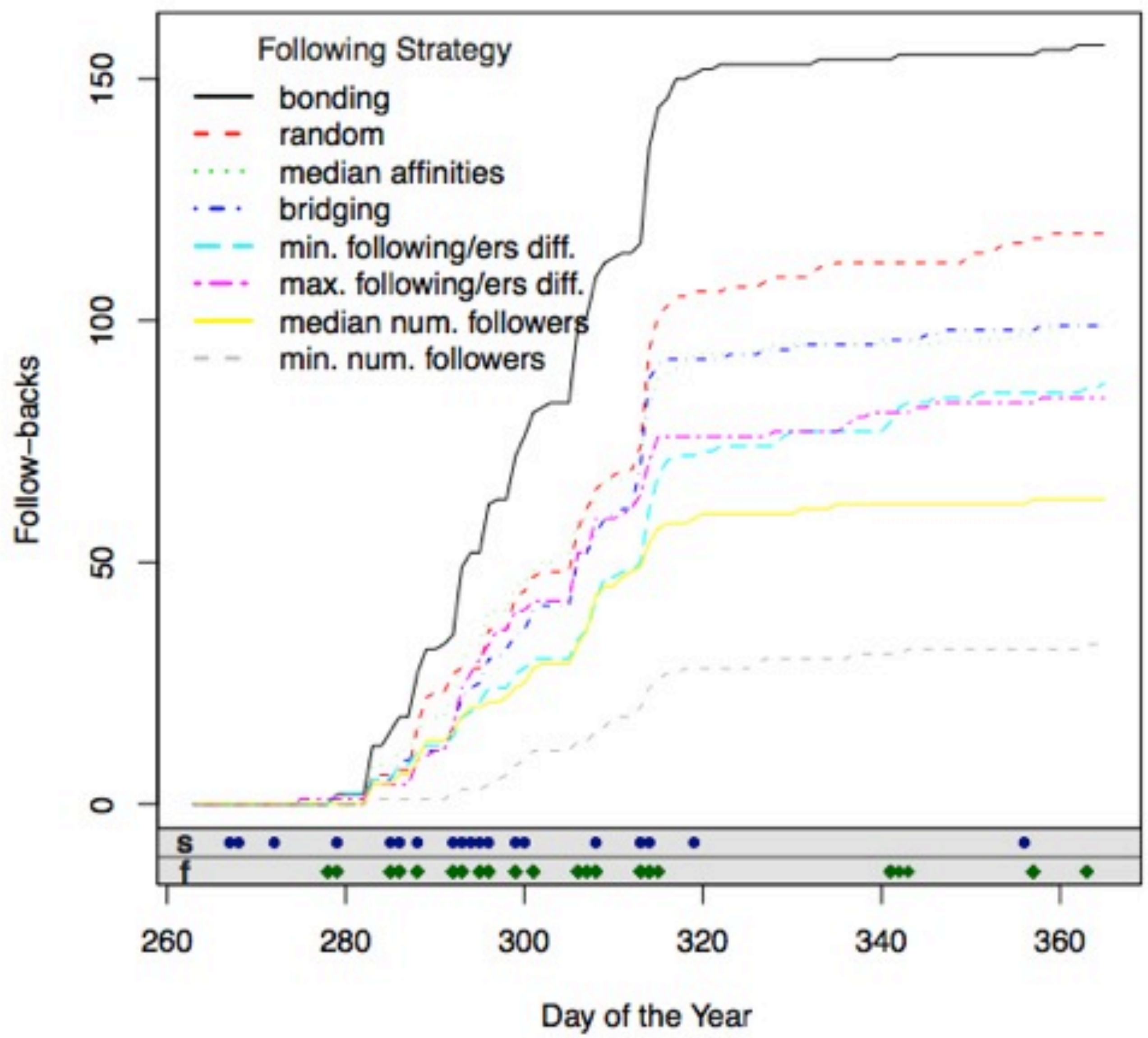


Fig. 2. **Follow-backs Over Time**: Follow-backs obtained by accounts in $\mathcal{A}$ throughout the duration of the study. Days in which following rounds occurred (i.e., accounts in $\mathcal{A}$ followed users in $\mathcal{U}$) are marked in the row labeled **f**. Days that new status updates were posted to the accounts in $\mathcal{A}$ are marked in the row labeled **s**.

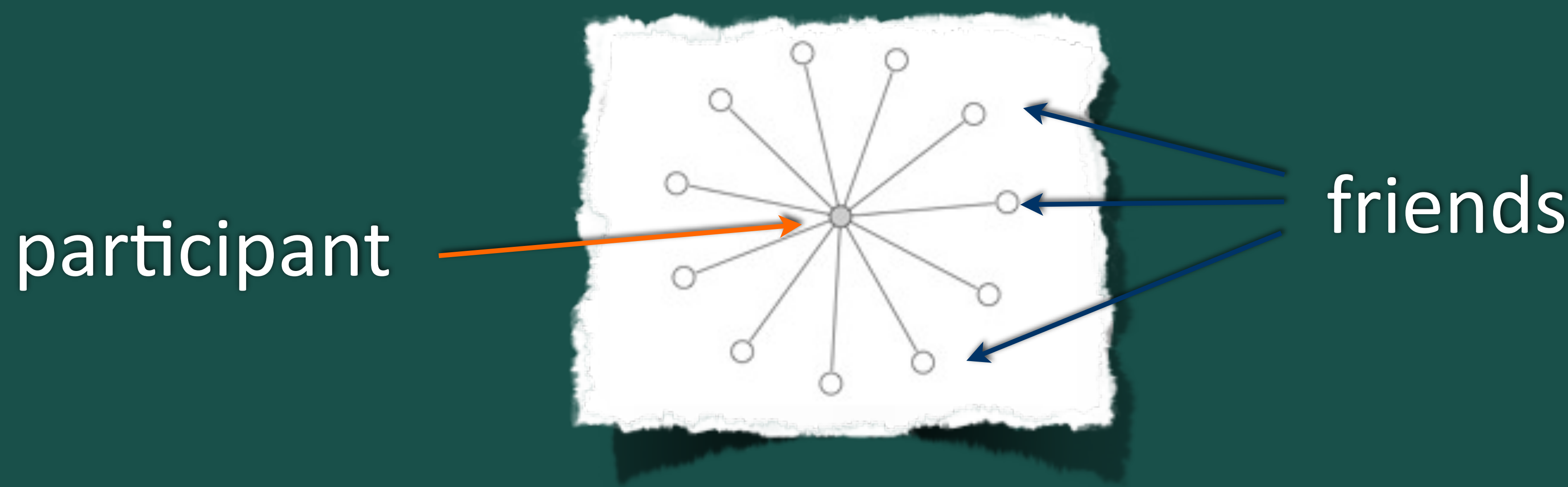
Language Acquisition

Experimental Setup

- Study abroad in Japan
- 204 participants
- Spent an average of 8.4 months in Japan, taking 13.2 hours per week of Japanese language courses in 38 language programs across 22 different cities
- Language improvement measured (pre-test and post-test)
- Offline community

Experiment Details

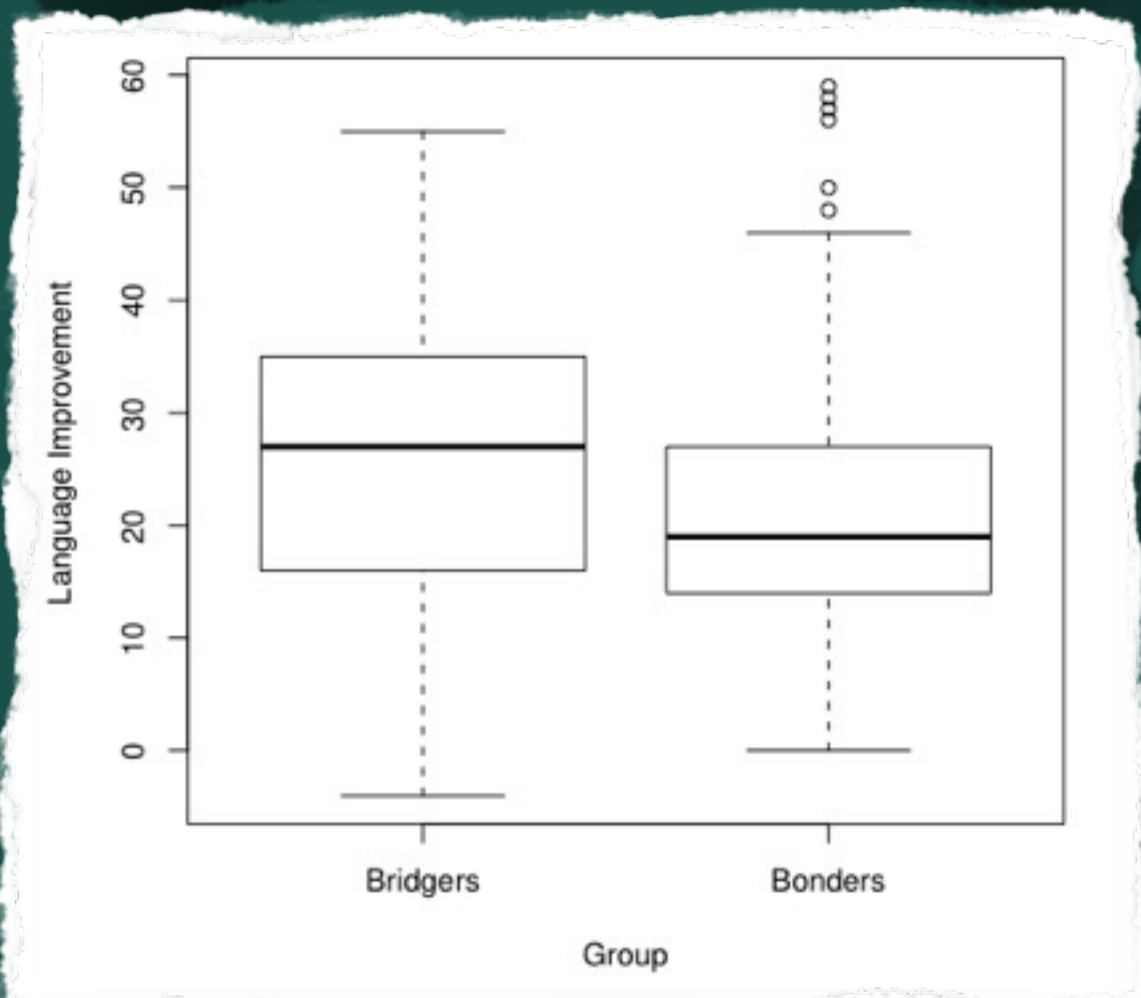
- Study abroad social interaction questionnaire (SASIQ)
- Friends and topics about which they spoke were listed
- Bonding and bridging scores were computed

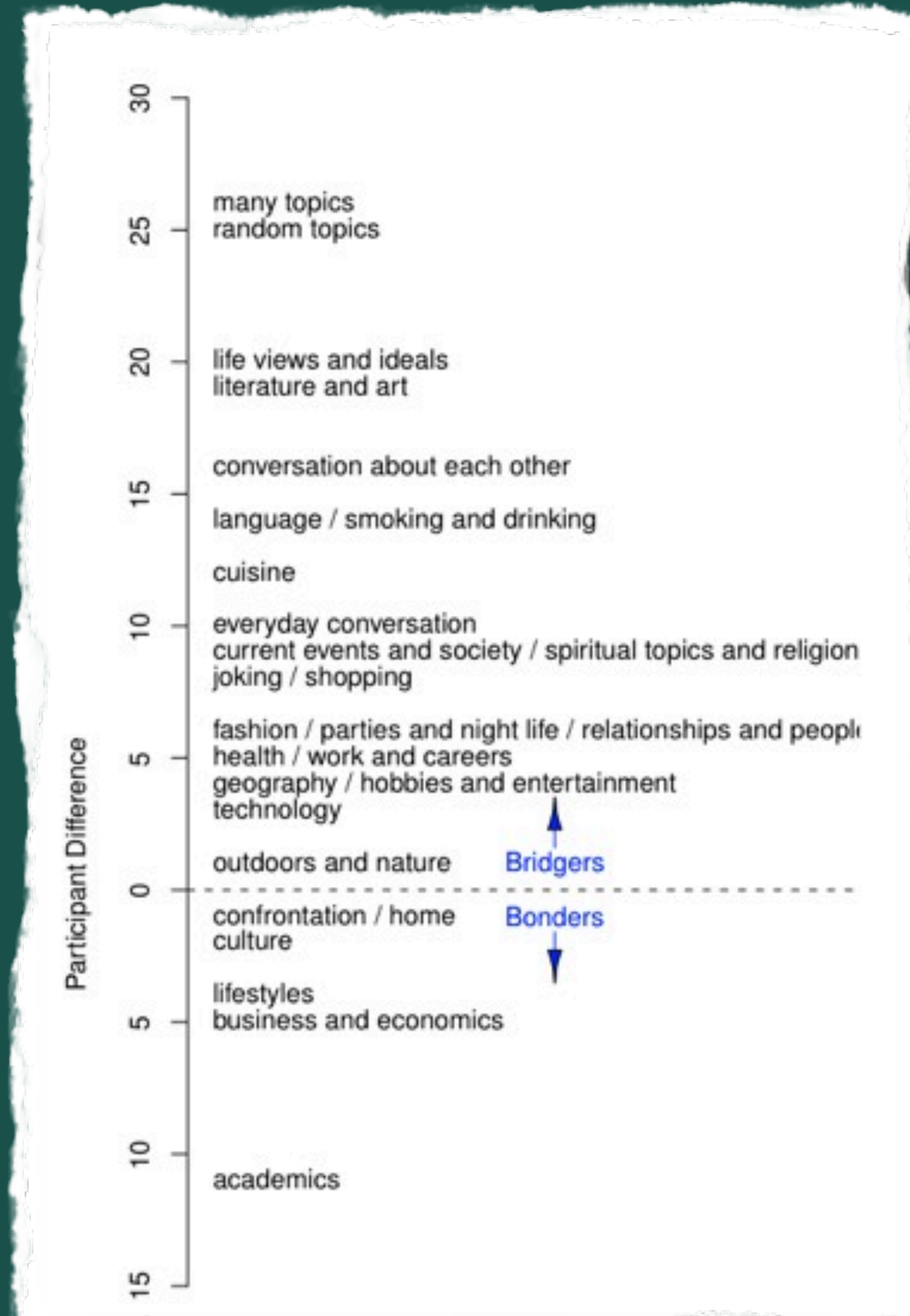


Results

Table 1: **Influence of Social Capital on Conversation:** Bridgers discuss more topics on average than bonders.

Group	Topics	Participants	Topics / Participant
Bonders	786	114	6.89
Bridgers	992	90	11.02





Publications (Case Studies)

- Social Capital in the Blogosphere: A Case Study (published in *Papers from the AAAI Spring Symposium on Social Information Processing*)
- Bonding vs. Bridging Social Capital: A Case Study in Twitter (published in *International Symposium on Social Intelligence and Networking*)
- Identifying Health-Related Topics on Twitter (published at *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction*)
- Social Capital and Language Acquisition during Study Abroad (published in *The 33rd Annual Conference of the Cognitive Science Society*)

Conclusion

- Case Study Contributions
  - Blogosphere - Highlighted potential bonding/bridging
  - Twitter - Tested and confirmed principle of homophily
  - Public Health - Increased understanding of tobacco-related behavior in Twitter
  - Language Acquisition - Bridging relationships achieve higher levels of language improvement

Conclusion

- **Computational Framework for Social Capital**
  - Unifies multiple definitions of social capital into general framework
  - Uses relationships, affinities, and resources
  - Hybrid networks - distinguishing potential and realized social capital
  - Bonding and bridging metrics vary independently (Putnam)
  - Supports *access* to and *mobilization* of resources
  - Notion of *interaction* formalized
  - Offers bridge between resource-sharing sites and social networks

Comparison of SCF

to Popular Applications

Feature	FB	TW	LI	FC	SW	SCF
relationships						
simple (friend or not)	✓	✓	✓			✓
complex (strength)						✓
asymmetric		✓				✓
dynamic						✓
affinities						
simple (like or not)	✓		✓		✓	✓
complex (how much)						✓
dynamic						✓
resources						
desired by ego				✓	✓	✓
desired by others				✓		✓
available by ego				✓	✓	✓
...willing to give				✓	✓	✓
...to specific others						✓
available by others				✓	✓	✓
...willing to give				✓	✓	✓
...to ego						✓
dynamic rel. (on exchange)						✓
...update on giving						✓
...update on receiving						✓

Future Work

- Perform additional resource simulations with multiple types of individuals within the network (e.g., producers and consumers, altruists and free-riders).
- Map practical selection functions to specific tasks (e.g., finding a job, obtaining support, learning a new skill).
- Utilize prior social exchange theory to seed reasonable hypotheses to then be empirically tested within the framework. Perhaps, suitable individual-level exchange functions could be developed.
- Design a range of relationship strength updating functions that allow external feedback (e.g., holidays, weather, crisis) to affect changes.

Future Work

- Explore the economics of automatically adjusting resource values based on the supply and demand within the network.
- Test additional well-developed theories of the social sciences to confirm validity within the context of specific online communities (e.g., can bonding social capital on Twitter be leveraged to lose weight?)
- Observe and track socially connected communities that actively exchange resources.
- Many of these ideas are finding their way into local startups (e.g., kalood, stubtopia)