



# The Early Stages of Social Capital: A Twitter Case Study

Matthew Smith

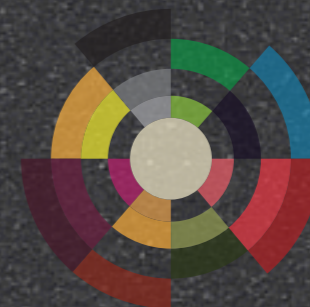
<http://m.smithworx.com>

Christophe Giraud-Carrier

[cgc@cs.byu.edu](mailto:cgc@cs.byu.edu)

<http://dml.cs.byu.edu>

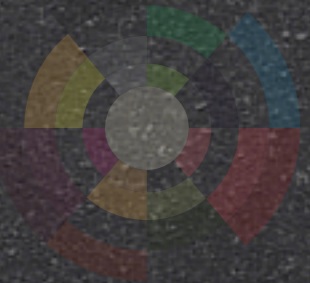
March 2010



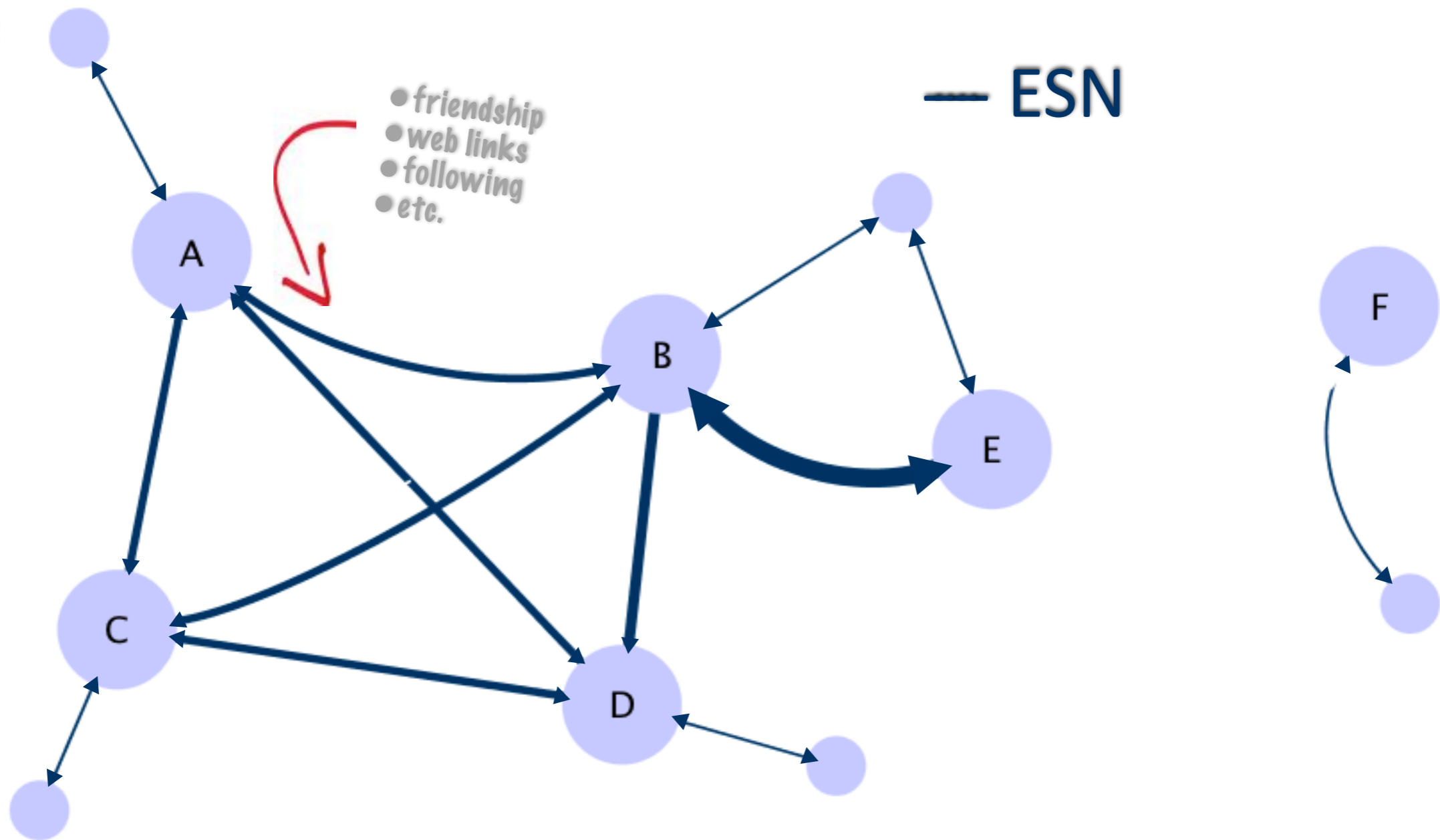
DML  
BYU DATA MINING LAB

# Introduction

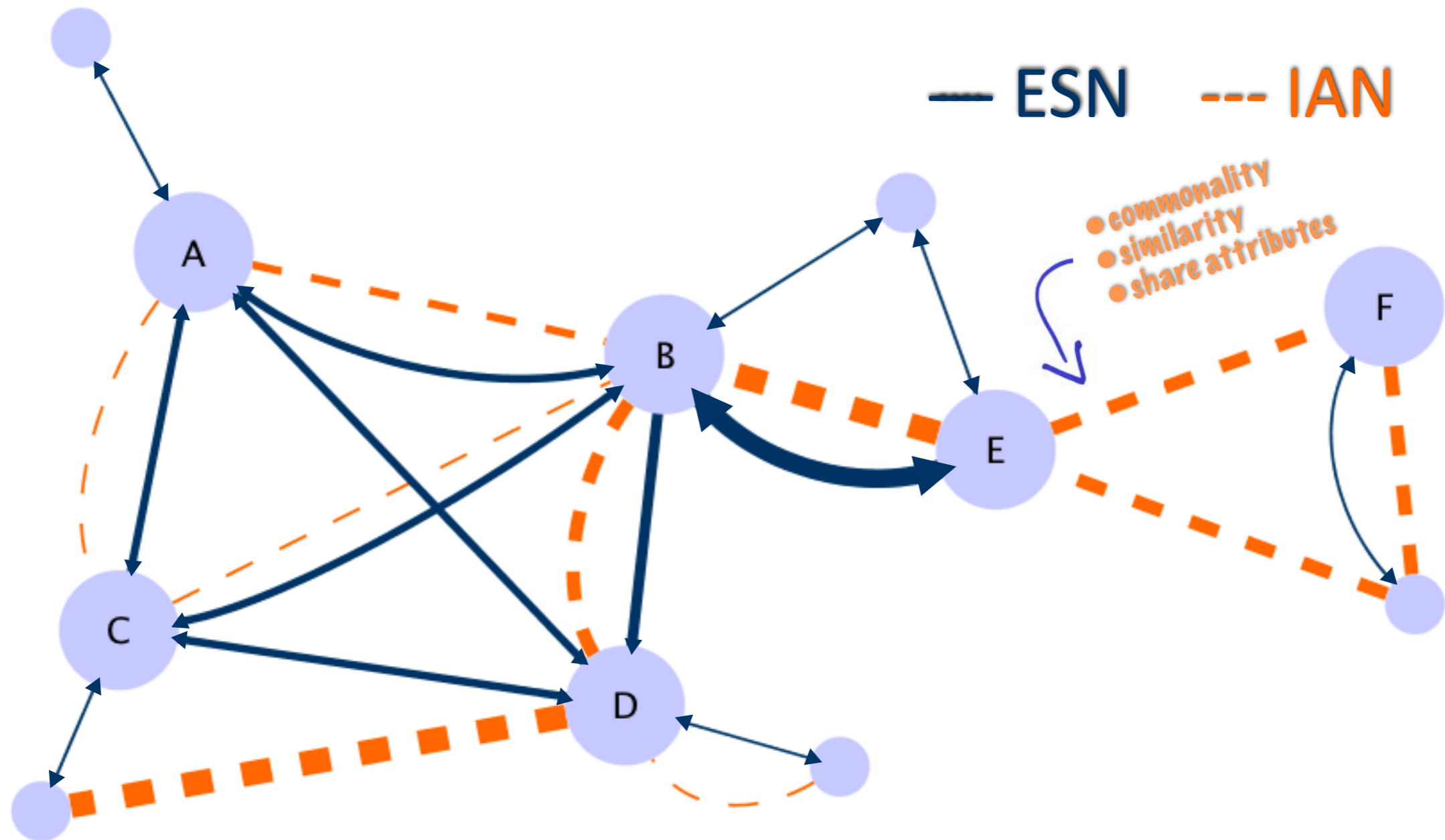
- Social Capital - leverage our online social networks
  - Develop framework that quantifies the value
  - Allows for theories to be tested
- This paper focuses on the following (in Twitter):
  - “Bonding is more likely to occur than bridging” - Lin 01
  - “Closure is the most obvious force” - Burt 05
  - Homophily principle: “Birds of a feather flock together”  
Smith-Lovin 87 “Similarity begets friendship”



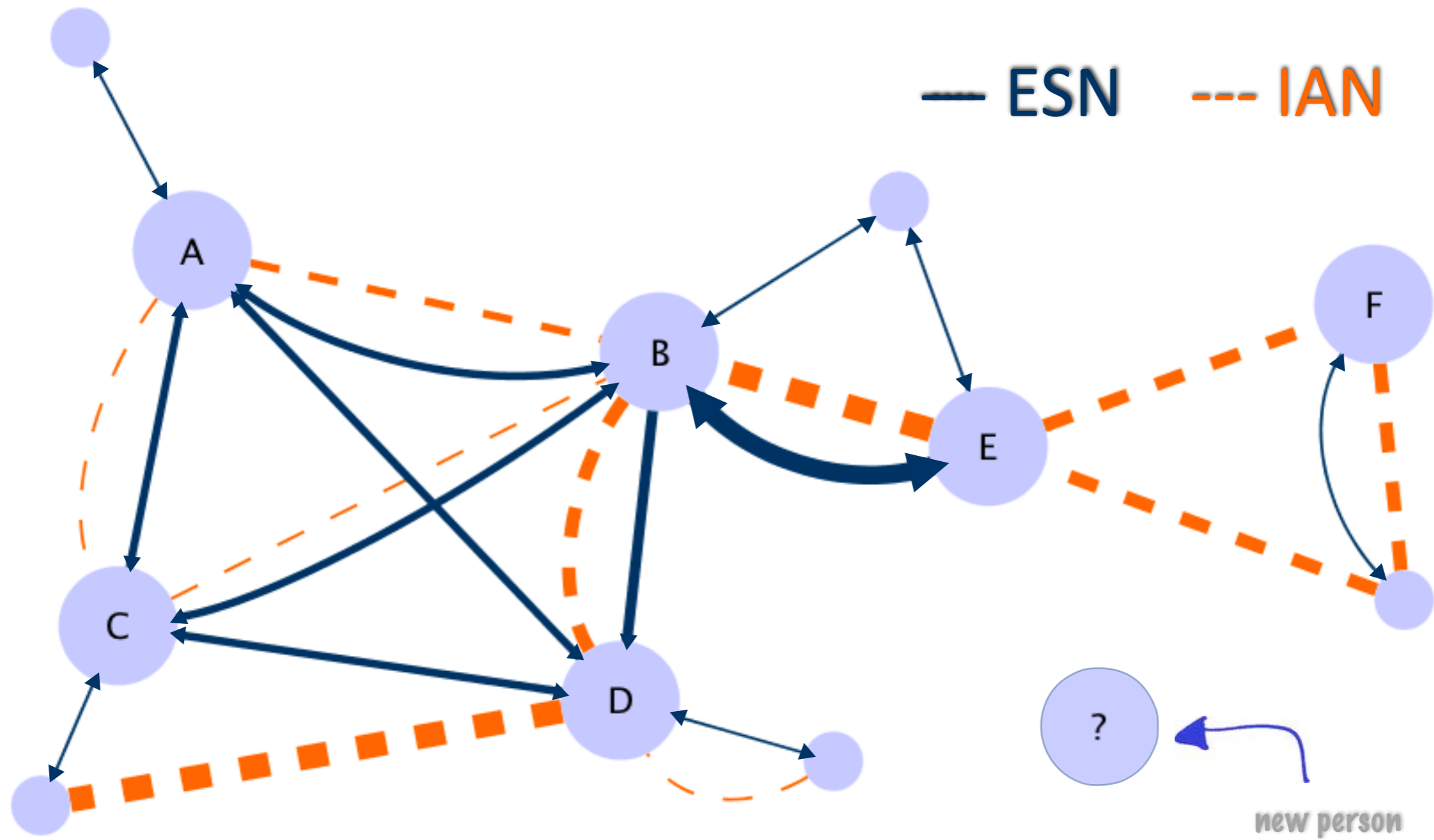
# Motivation



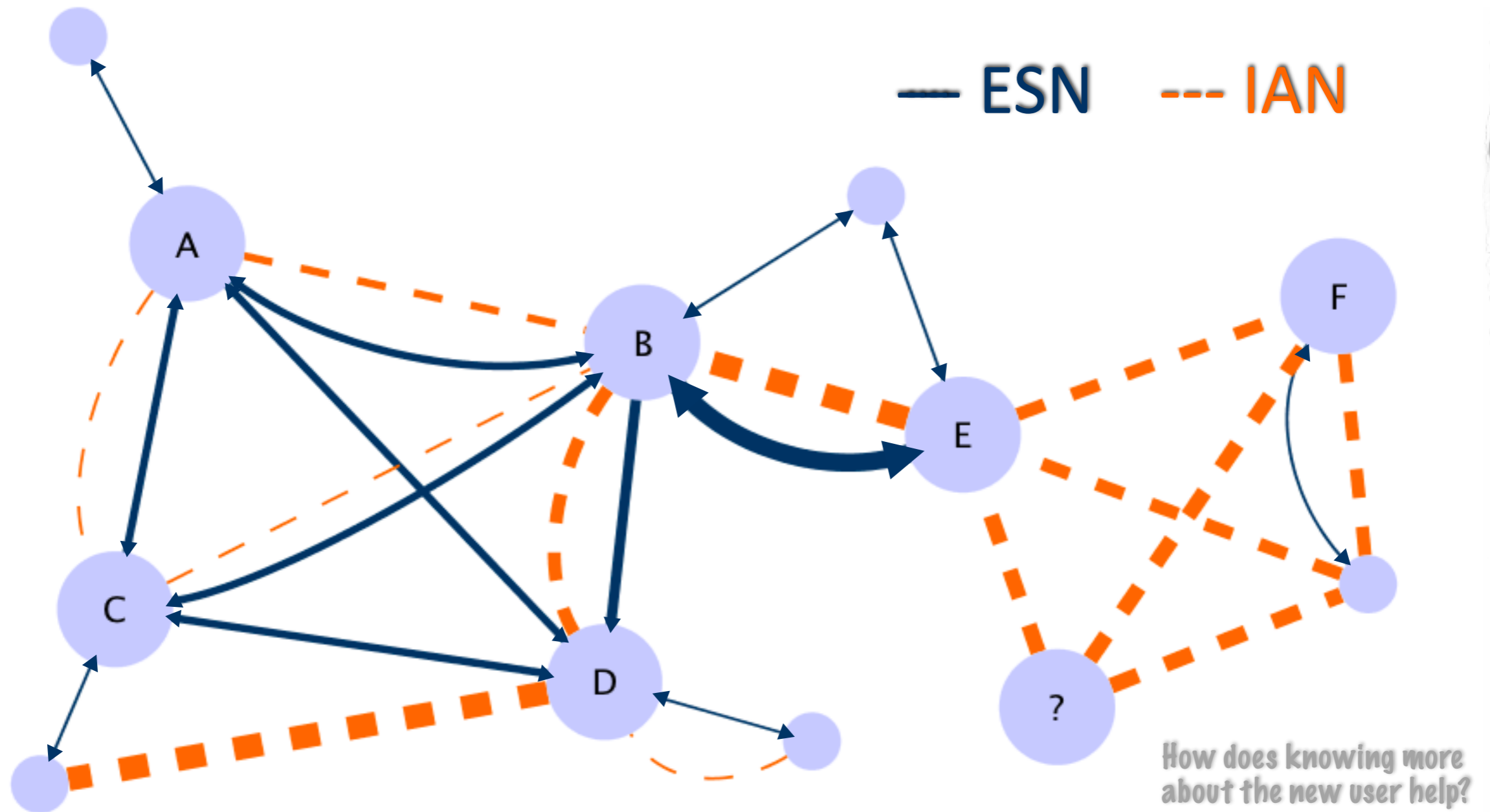
# Motivation



# Motivation



# Motivation



# Twitter Basics - Profile

The screenshot shows a Twitter profile for 'smithworx' (Matt Smith). The profile includes a header with navigation links (Home, Profile, Find People, Settings, Help, Sign out), a profile picture, name, location, website, and bio. The main content area shows a 'That's you!' section with a 'Lists' dropdown and a 'my tweets (reverse chronological order)' section containing four tweets. The right sidebar shows statistics (493 following, 621 followers, 19 listed), a 'Tweets' count (1,200), a 'Favorites' section, a 'Following' list of user avatars, and an RSS feed link.

**twitter** Home Profile Find People Settings Help Sign out

 **smithworx** *link >*

Name Matt Smith  
Location Utah  
Web <http://m.smithworx.com>  
Bio social capital guru, web startup and new venture man, phd candidate, founder @webstartupgroup

**493** **621** **19**  
following followers listed

**Tweets** 1,200

**Favorites**

**Following**

*stats >*

**That's you!** Lists ▾

Your lists: [dml](#)

*my tweets (reverse chronological order)*

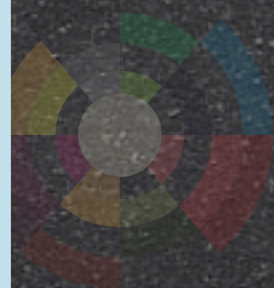
**WebStartupGroup** StartupNews: The Rules of Extreme Programming <http://bit.ly/cQp1yl>  
about 2 hours ago via twtsimple  
Retweeted by you

**emilam** Hey Instructure BYU is reassessing blackboard this month. Can you please win their love  
about 5 hours ago via TweetDeck  
Retweeted by you

Preparing my presentation for the Spring Research Conference this weekend.  
about 6 hours ago via web

**@mlefrandt** I sent you a couple emails last week --- no worries. I'm glad to hear that lots is going on with the social startup scene.  
about 22 hours ago via web in reply to mlefrandt

RSS feed of smithworx's tweets



# Twitter Basics - Home

The screenshot shows the Twitter home page. At the top left is the Twitter logo. To the right is a navigation bar with links for Home, Profile, Find People, Settings, Help, and Sign out. Below the navigation bar is a text input field with the placeholder "What's happening?" and a character count of "140". Below the input field is a "Tweet" button. A notification for a retweet is visible: "Latest: WebStartupGroup StartupNews: The Rules of Extreme Programming http://bit.ly/cQp1yl Retweeted by you 8 minutes ago". Below the input field is a green box with the text "tweets (by those I follow)". The main content area shows a list of tweets. The first tweet is from CNN: "Haitian children taken by American missionaries reunite with families http://on.cnn.com/cLoyWM half a minute ago via web". The second tweet is from larrybrauner: "30 Tips for Using Social Media in Your Business http://ping.fm/5GHdH half a minute ago via Ping.fm". The third tweet is from ashbuckles: "#chromeitup 1 minutes ago via TweetDeck". The fourth tweet is from larrybrauner: "How to Use Social Networking Sites to Drive Business http://amplify.com/u/3bxi 3 minutes ago via Amplify". The fifth tweet is from WebStartupGroup: "New Blog Post: yConnect: March 25, 2010 http://bit.ly/b0fKbK 3 minutes ago via twtsimple". On the right side, there is a profile card for smithworx with 1,200 tweets, 493 following, 621 followers, and 19 listed. Below the profile card is a "Chirp" section with the text "n. Twitter's first developer conference in April." Below the chirp is a "Home" section with the text "@smithworx".

twitter

Home Profile Find People Settings Help Sign out

What's happening? 140

Latest: WebStartupGroup StartupNews: The Rules of Extreme Programming <http://bit.ly/cQp1yl> Retweeted by you 8 minutes ago Tweet

tweets  
(by those I follow)

Home

**cnnbrk** Haitian children taken by American missionaries reunite with families <http://on.cnn.com/cLoyWM> half a minute ago via web

**larrybrauner** 30 Tips for Using Social Media in Your Business <http://ping.fm/5GHdH> half a minute ago via Ping.fm

**ashbuckles** #chromeitup 1 minutes ago via TweetDeck

**larrybrauner** How to Use Social Networking Sites to Drive Business <http://amplify.com/u/3bxi> 3 minutes ago via Amplify

**WebStartupGroup** New Blog Post: yConnect: March 25, 2010 <http://bit.ly/b0fKbK> 3 minutes ago via twtsimple

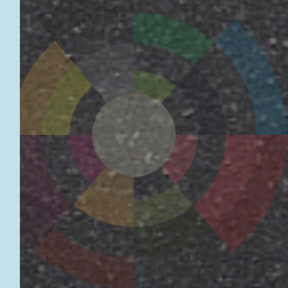
**smithworx**  
1,200 tweets

493 following 621 followers 19 listed

**Chirp**  
n. Twitter's first developer conference in April.

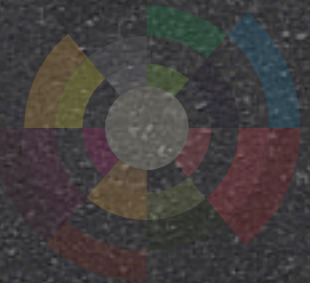
Home

@smithworx



# Hypothesis

*Utilizing a following strategy motivated by a desire to maximize potential bonding social capital produces higher returns (or social success) than utilizing other strategies.*



# Experimental Setup

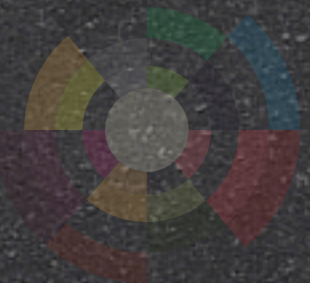


- Let  $\mathcal{A}$  be the set of new Twitter accounts ( $|\mathcal{A}|=9$ )
  - Assign each a following strategy
  - Each behave identically (except for their following strategy)
- Let  $\mathcal{U}$  be the set of sampled Twitter users to follow ( $|\mathcal{U}|=5000$ )
  - tweeted “data mining” or “social networks” in Sept. '09
- Compare stats for each strategy; all accounts in  $\mathcal{A}$ 
  - Test if strategy  $A$  (*max. pot. bonding*) results in higher returns



# Experimental Setup - Strategies

- (A) max. potential bonding
- (B) max. potential bridging
- (C) median affinities
- (D) random
- (E) min. following/ers diff.
- (F) max. following/ers diff.
- (G) median num. followers
- (H) min. num. followers



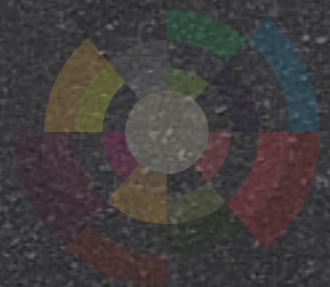
# Experimental Setup - Selection

$u$ :  
Pool of 5000



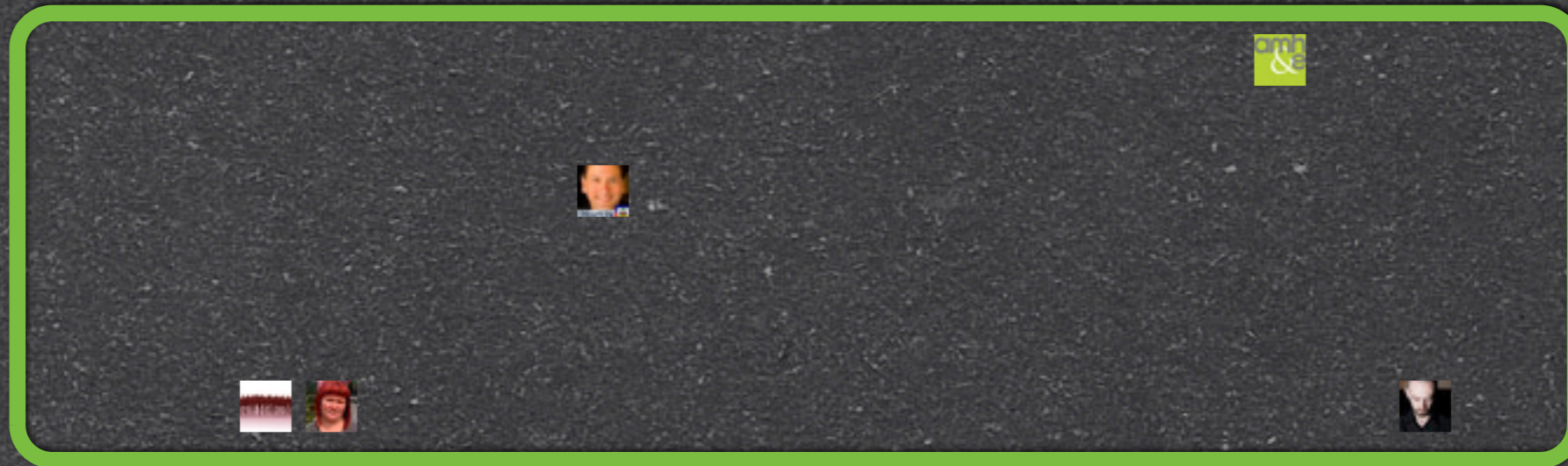
Begin Selection...

$a$ :



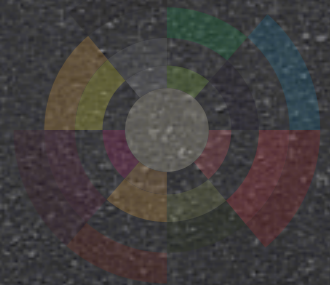
# Experimental Setup - Selection

$u$ :



Stop When Each Account  
is Following 500 Users

$a$ :



# Experimental Setup - Tracking



**Jon the Magnificent**

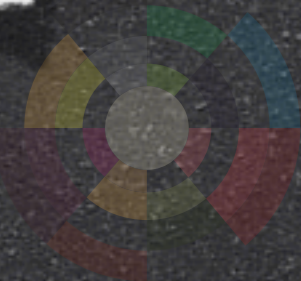
Hi, I'm Jon. Thanks for visiting my homepage! I love the Internet, Twitter, and life. If you would like to contact me, you can send me a message on Twitter or send email to: [jon at twittercapital.com](mailto:jon@twittercapital.com)



# Experimental Results

rank <sub>f</sub>	strategy	following	follow-backs	↓ followers	rejects	churn	follow <sub>total</sub>	follower <sub>total</sub>
1	<i>max. potential bonding</i> (A)	500	158 (32%)	202 (40%)	12	127	512	329
2	<i>max. following/ers diff.</i> (F)	500	84 (17%)	172 (34%)	12	324	512	496
3	<i>random</i> (D)	500	118 (24%)	154 (31%)	20	103	520	257
4	<i>median affinities</i> (C)	500	99 (20%)	123 (25%)	25	93	525	216
5	<i>max. potential bridging</i> (B)	500	99 (20%)	120 (24%)	25	91	525	211
6	<i>min. following/ers diff.</i> (E)	500	87 (17%)	99 (20%)	50	55	550	154
7	<i>median num. followers</i> (G)	500	63 (13%)	86 (17%)	31	51	531	137
8	<i>min. num. followers</i> (H)	500	33 (07%)	42 (08%)	79	29	579	71
9	<i>follow nobody</i> (I)	0	0 (—%)	3 (—%)	0	24	0	27

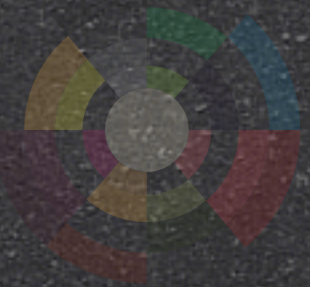
Table 1: **Follower Statistics:** Each of the nine accounts are listed by *strategy* and ranked by the number of *followers* obtained during the experiment, denoted *rank<sub>f</sub>*. The *following* column is the number of users that the account was following at the end of the experiment. The *follow-backs* column reports the number of users followed that were following the account back at the end of the study, the percent of following is supplied for reference (i.e., *follow-backs/following*). The *followers* column is the number of users following the account at the end of the experiment (including those that were never followed by the account). The percent of following is also supplied for reference (i.e., *followers/following*). The *rejects* column reports the number of users that could not be followed on Twitter at the time (e.g., account was protected, user attempting to follow was blocked, or user was suspended). The *churn* statistic reports the number of users that followed the account for a time, but were no longer following the account at the end of the experiment. The *follow<sub>total</sub>* is the total number of users that were followed by the account, i.e., the sum of *following* and *rejects*. The *follower<sub>total</sub>* is the total number of users that followed the account during the experiment, i.e., the sum of *followers* and *churn*.



# Results: Followback-to-Following

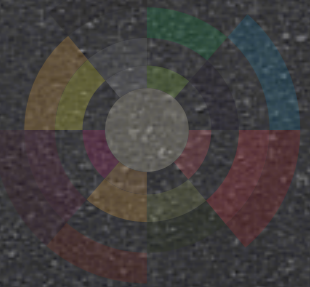
strategy	significantly different
(A) <i>max. potential bonding</i>	B, C, E, F, G, H
(B) <i>max. potential bridging</i>	A, H
(C) <i>median affinities</i>	A, G, H
(D) <i>random</i>	G, H
(E) <i>min. following/ers diff.</i>	A, H
(F) <i>max. following/ers diff.</i>	A, H
(G) <i>median num. followers</i>	A,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

Table 3: **Followback-to-Following:** Pairwise Proportion Test Results. ( $\alpha = 0.01$ , Bonferroni corrected  $p$ -values)



# Results - Clicks & Mentions

$\text{rank}_c$	$\text{rank}_f$	strategy	$\text{clicks}_t$	$\text{clicks}_p$	↓ total clicks	mentions
1	3	<i>random</i> (D)	900	9	909	2
2	1	<i>max. potential bonding</i> (A)	882	15	897	3
3	8	<i>min. num. followers</i> (H)	850	16	866	1
4	2	<i>max. following/ers diff.</i> (F)	849	7	856	1
5	4	<i>median affinities</i> (C)	846	9	855	1
6	6	<i>min. following/ers diff.</i> (E)	821	19	840	4
7	5	<i>max. potential bridging</i> (B)	773	11	784	2
8	9	<i>follow nobody</i> (I)	775	1	776	1



# Results - Clicks vs. Followers

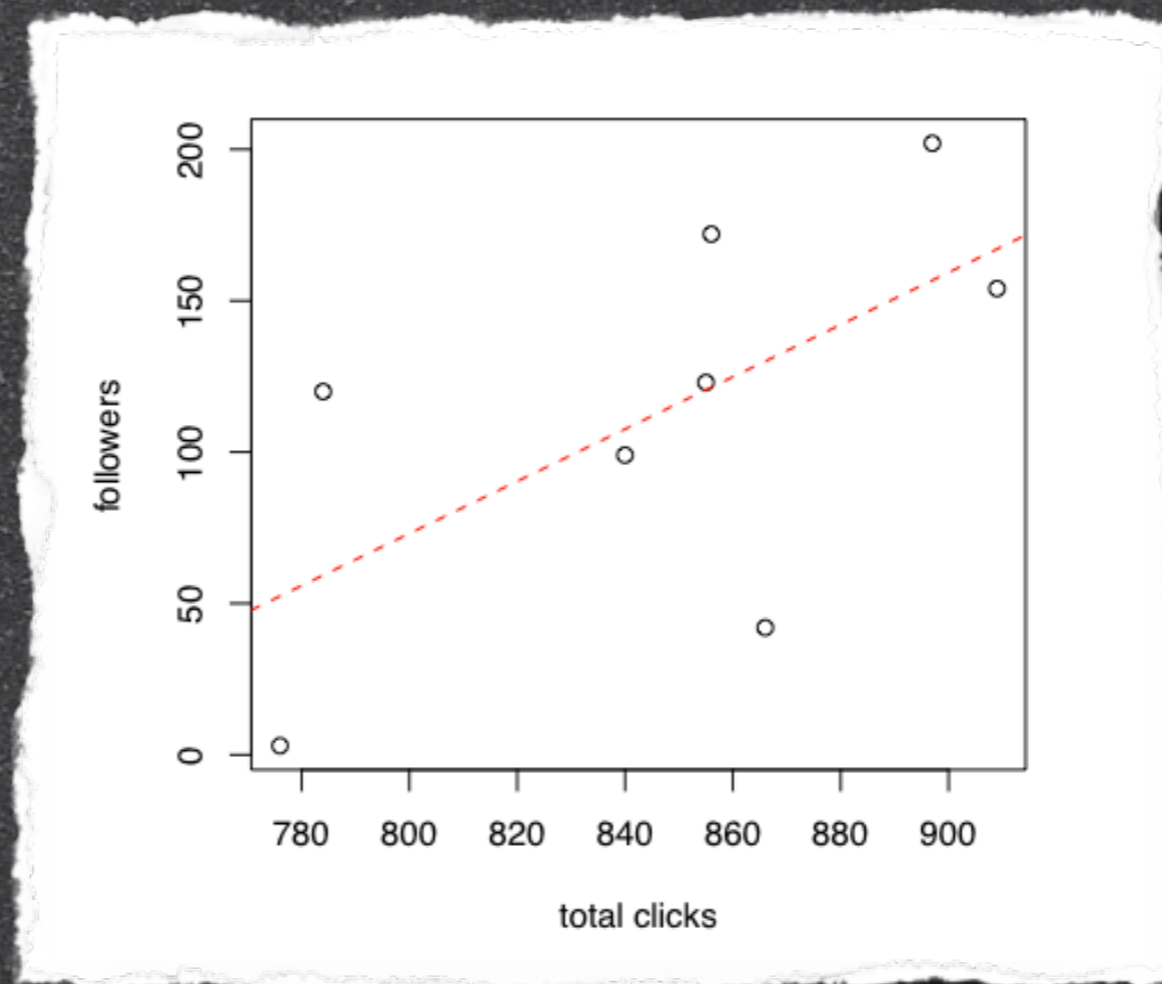
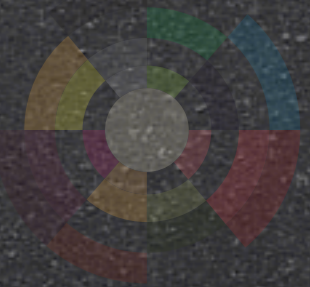


Figure 3: **Clicks vs. Followers.** The linear model shown by the regression line (dashed) poorly fits the data having an  $R^2$  value of 0.28. There is, however, a positive Pearson correlation of 0.62, yet it is not as high as might be expected.



# Results - Follow-backs Retained

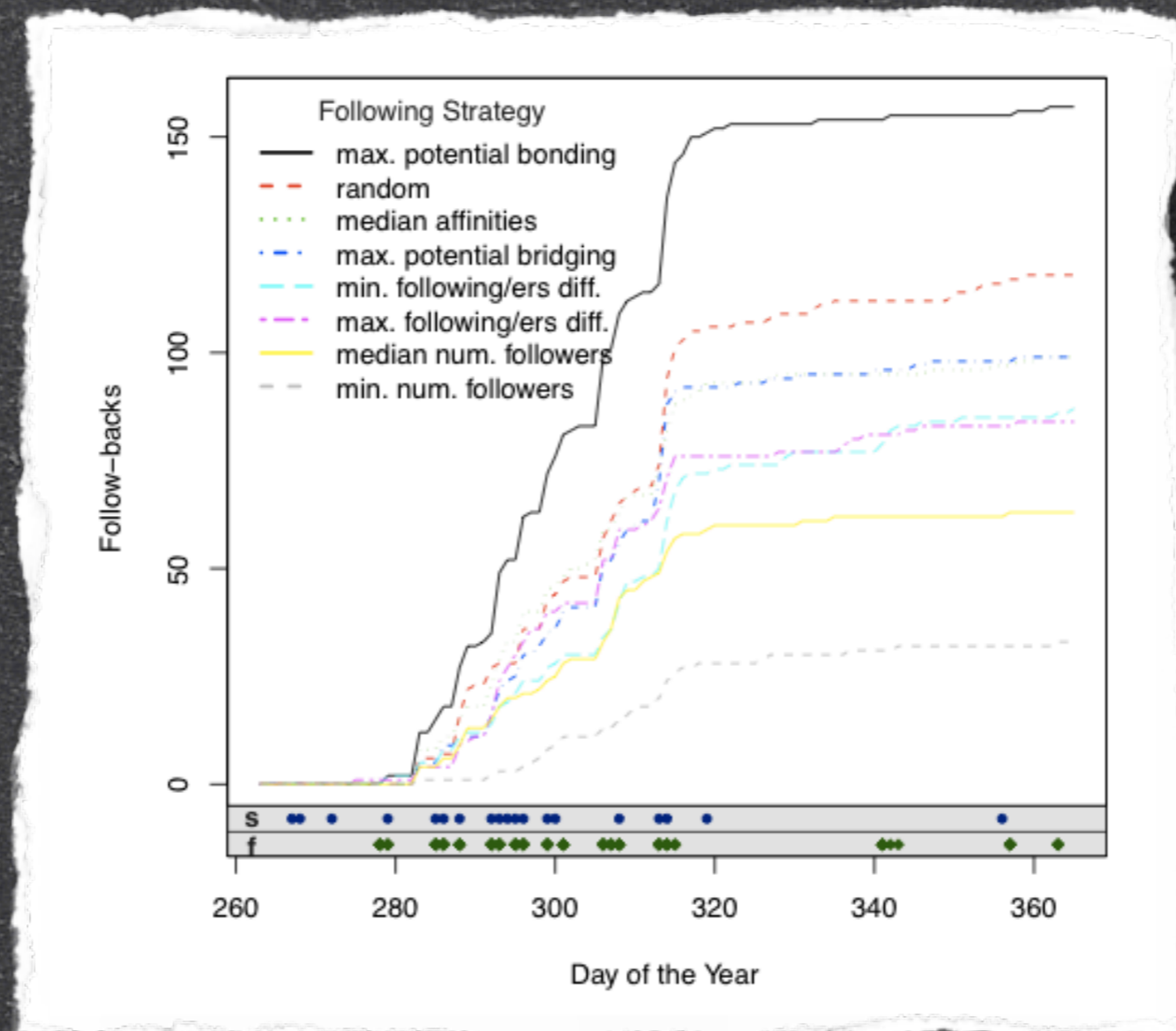
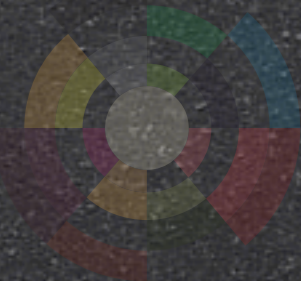
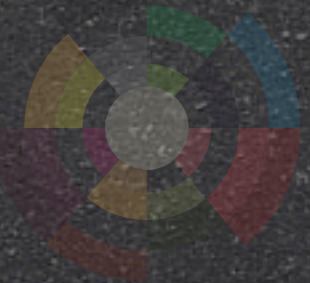


Figure 3: **Follow-backs Over Time:** Follow-backs obtained by accounts in  $\mathcal{A}$  throughout the duration of the study. Days in which following rounds occurred (i.e., accounts in  $\mathcal{A}$  followed users in  $\mathcal{U}$ ) are marked in the row labeled **f**. Days that new status updates were posted to the accounts in  $\mathcal{A}$  are marked in the row labeled **s**.



# Conclusions

- New users that requested to follow others having similar profile descriptions (i.e., maximize potential bonding) significantly increased the number of follow-backs
- This empirical evidence confirms the prevalent sociological beliefs
- This research assists new users to determine who to connect to in order to grow their network



# Questions & Comments

Ask me now:



Email or Call me:

**Matt Smith**  
**smitty@byu.edu**  
**(801) 788-4920**

---

Connect:

**Web:** <http://m.smithworx.com>

**Blog:** <http://dmine.blogspot.com>

**Facebook:** <http://facebook.com/smithworx>

**LinkedIn:** <http://linkedin.com/in/smitty>

**Twitter:** <http://twitter.com/smithworx>

