

Classifying the Political Leaning of News Articles and Users From User Votes

Xiaodan Daniel Zhou

Paul Resnick

Qiaozhu Mei

[Gadgets](#)[Themes](#)

US Political News & Opinion

- [Caroline Kennedy to Seek Clinton's ...](#)
- [Limits on Executive Pay May Prove Toothle...](#)
- [Washington Wire - WSJ.com The Real Story...](#)
- [Rasmussen Reports™ The most compreh...](#)
- [Google Public Policy Blog Net neutrality ...](#)
- [Kast sko mot Bush](#)
- [Jules Crittenden » The Bulge](#)
- [Brother Explains Shoe-Tossing Iraqi Journ...](#)
- [States' Funds for Jobless Are Dryin...](#)
- [Iraqis Hail Journalist Who Threw Shoes at...](#)
- [Think Progress » Iraqi Journalis...](#)
- [The Page - by Mark Halperin - TIME](#)

US Political News & Opinion

[« Back to gallery](#)

By [Sean Munson](#)

Top US political news and opinion stories from the last 48 hours, based on links from over 500 blogs. Links are highlighted to show the average political leaning of the blogs that are talking about each item, but you can disable this in the settings.

Rectangular Snip

< 100 users | ★★★★★ 0 ratings

Add it now

 [Share this gadget](#)

<http://tinyurl.com/dy53jq>

Most Recent

Top in 24 Hours

7 Days

30 Days

Submit a link

Digg it ↑

All Topics

Business

Entertainment

Gaming

Lifestyle

Offbeat

Politics*

Science

Sports

Technology

World News*

Politics News

AI

Images

Videos

Stories

172

Why Julian Assange Is 2010's Person Of The Year



prosebeforehos.com — For striking fear into the heart of every crooked politician and banker and for showing us the gift—and the curse—of too much information, Julian A... 1 hr 32 min ago

via rudymoman

36 Comments

Save

Bury

208

Ellsberg: "EVERY attack now made on WikiLeaks and Julian Assange was made against me and the rele...



ellsberg.net — 7 hr 42 min ago

via jamesg546

36 Comments

Save

Bury

145

VIDEO: Maddow And Olbermann Slam Obama's Tax Cut Compromise

tpmdc.talkingpointsmemo.com — If you thought you were mad about: Obama's tax cut compromise, you should watch Keith Olbermann and Rachel Maddow slam it. 7 hr 12 min ago

via worketiemegan

147 Comments

Save

Bury

today's freebie

balance man

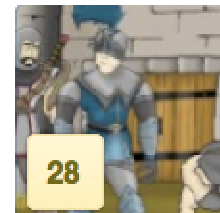
about

You are perfectly balanced

my history



Hot Stories on Digg



28

HOT STORY

The Escapist : Video Galleries : Zero Punctuation : Assassin's Creed: Brotherhood

escapistmagazine.com

Chosen by digg staff. [Learn more.](#)

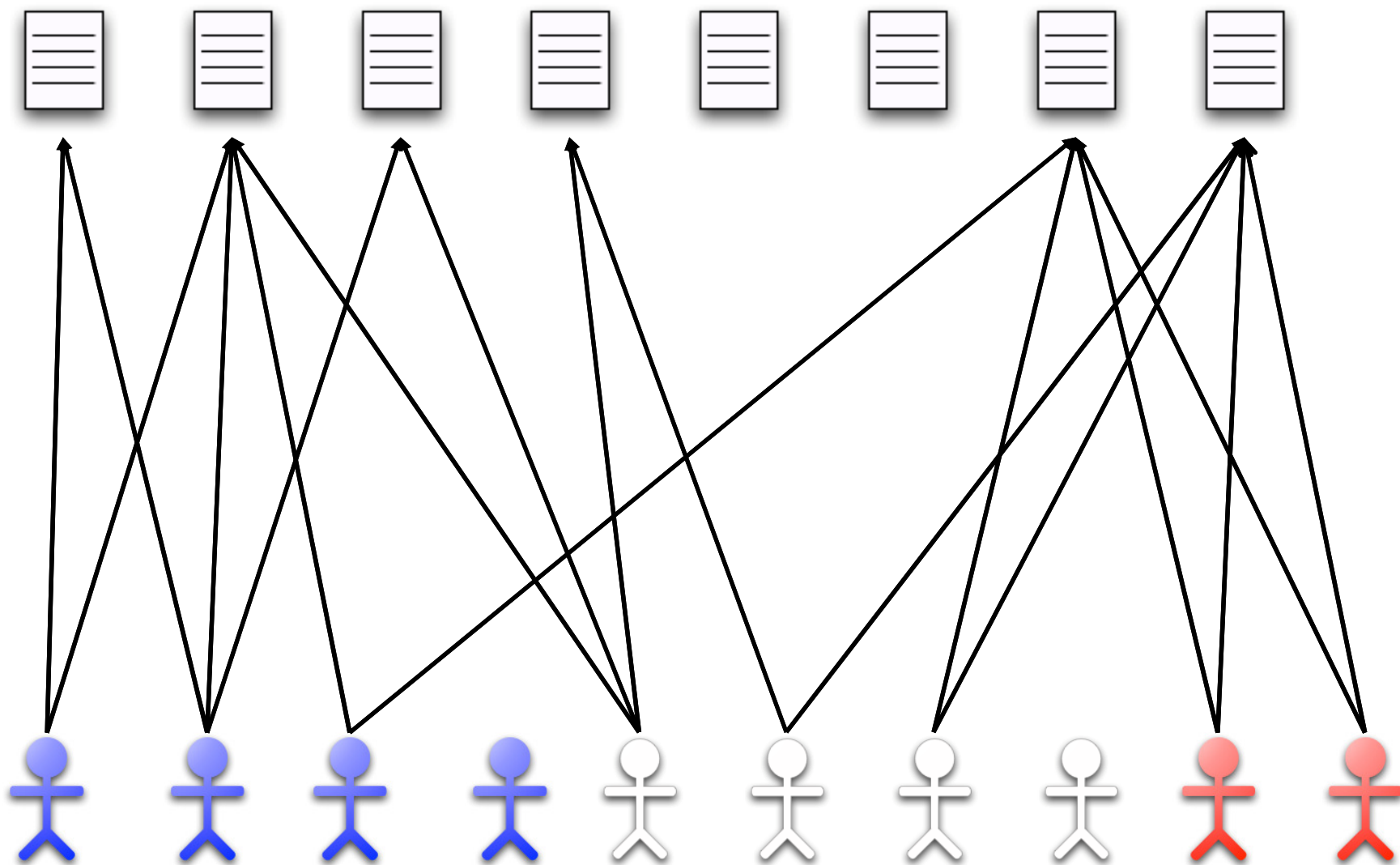
[Give us feedback](#) | [Take the survey](#)

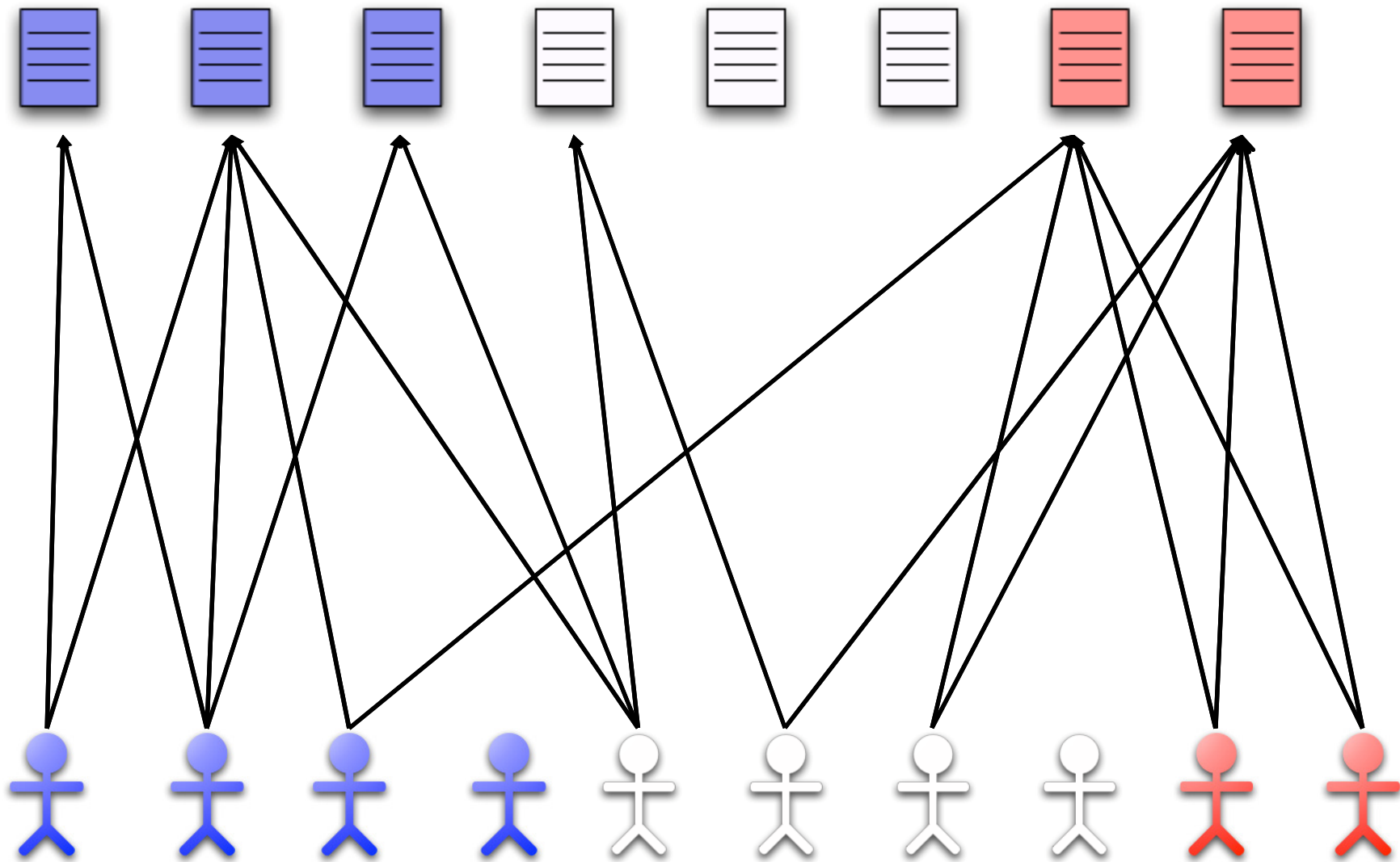
Text Classification

- (Oh et al, ICWSM 2009)
- SVM
- 87% accuracy on blog posts









digg

Digg - Politics - The Latest

+

←

→

↻

digg.com/news/politics

☆

1

digg™

My News

Top News

Upcoming

0 Followers

0 Following

Most Recent

Top in 24 Hours

7 Days

30 Days

Submit a link

Digg it ↑

All Topics

Business

Entertainment

Gaming

Lifestyle

Offbeat

Politics

Science

Sports

Technology

World News

Politics News

All

Images

Videos

Stories

82

Digg ↑

Did President Obama Receive an Illegal Contribution?

[blogs.browardpalmbeach.com](#) — Thanks to a disgraced politician in South Florida, President Obama's campaign may have unknowingly benefitted from a \$15,000 check. 37 min ago

via [jgregc](#)

18 Comments

Save

Bury

104

Digg ↑

Poll: Voters like Clinton-backed candidates more than Obama

[articles.latimes.com](#) — The poll found that 53% of Democrats said Clinton's personal endorsement would carry weight compared to 48% who said Obama's would. 1 hr 18 min ago

via [llammalane](#)

14 Comments

Save

Bury

56

Digg ↑

The 'Must Have' Gift For The Holidays

[now.sprint.com/holiday](#) — At under 1lb, learn why the Samsung Galaxy Tab with 7" touchscreen, dual cameras, and flash capabilities, is a must for your wish list this season.

Digg v4: 502.11ms

Give us feedback

Take the survey

digg

Digg - Llama Lane (llamma ✕)

←

→

↻

digg.com/llammalane/diggs

digg™


My News

Top News

Upcoming

0 Followers

0 Following



71

Followers

9

Following

1,576

Diggs

5

Daily Diggs

80

Llama Lane

AllDiggsSubmissionsCommentsFollowersFollowing

109

Digg ↑

Bizarre Avatar-themed weddings take place in China
nerve.com — Avatar-themed weddings in Wulingyuan, a scenic area in Hunan Province, China become popular. 1 day 7 hr ago
via meetjoewhite 1 Comment Save Bury

104

Digg ↑

2011 Looks Gloomy For Renewable Energy
naturalresourcereport.com — With the 2010 elections behind us, environmental experts worry that the political shift in the U.S. House of Representatives may reduce federal spe... 1 day 6 hr ago
via kimballo 7 Comments Save Bury

145

Digg ↑

A Drop of Water at 2,000 Frames Per Second
thatvideosite.com — Not what you would expect to see 1 day 20 hr ago
via wjappe 6 Comments Save Bury

104

The Ultimate Elisha Cuthbert Gallery
topcultured.com — Elisha Cuthbert is hot, no doubt about it. But how hot? Hot

Give us feedback

Take the survey

Digg v4: 232.76ms

Massive Censorship Of Digg

blogs.alternet.org/oleoleolson/2010/08/05/massive-ce...

546 COMMENTS

ARTICLE TOOLS

TEXT SIZE:
A A A

EMAIL
PRINT

SHARE:

3559 tweets
retweet


7074
f Share

submit to reddit

12K

Posted by [oleoleolson](#) at 4:40 am
August 5, 2010

Massive Censorship Of Digg Uncovered



A group of influential conservative members of the behemoth social media site Digg.com have just been caught red-handed in a widespread campaign of censorship, having multiple accounts, upvote padding, and deliberately trying to ban progressives. An undercover investigation has exposed this effort, which has been in action for more than one year.

DiggPatriots - Conservatives & Libertarians

 Search for other groups...

 Search

[Home](#)
[Attachments](#)
[Members Only](#)
[Messages](#)
[Post](#)
[Files](#)
[Photos](#)
[Links](#)
[Database](#)
[Polls](#)
[Members](#)
[Calendar](#)
[Groups Labs \(Beta\)](#)
[Info](#) [Settings](#)
[Group Information](#)
[Members: 84](#)
[Category: Politics](#)
[Founded: May 26, 2006](#)
[Language: English](#)
[!\[\]\(a2bb1e57b467f1e41142026aa73db90f_img.jpg\) Already a member? Sign in to Yahoo!](#)
[Yahoo! Groups Tips](#)
[What you know...](#)
[Meet people, find stories, find how
 new Yahoo! Groups impacts members worldwide.](#)
[Best of Y! Groups](#)
[Check them out and
 nominate your group.](#)

[Follow us on Twitter](#)
[Visit the Groups Blog](#) for the latest Yahoo! Groups information.

[Home](#)
[Join This Group!](#)

 Activity within 7 days: [7](#) New Links - [537](#) New Messages

Description

"We need to continue the fight...not just on Digg, but networking with other conservative sites and getting the word out.

Positive Attitude! We're conservatives and our philosophy is based on the concept that if a person works hard, keeps faith, a positive attitude and has or has near to the gratitude that anything is possible. Liberals are full of envy and the belief that someone owes them. They have no clue what responsibility or independence is. They've been brainwashed into thinking that they are independent thinkers because they don't subscribe to a faith (an offense to any agnostic or atheist among us) while subscribing to faith in government and they don't even realize it. Their religion is governmental stupidity and they don't realize their philosophy really sucks a\$\$."



Message History

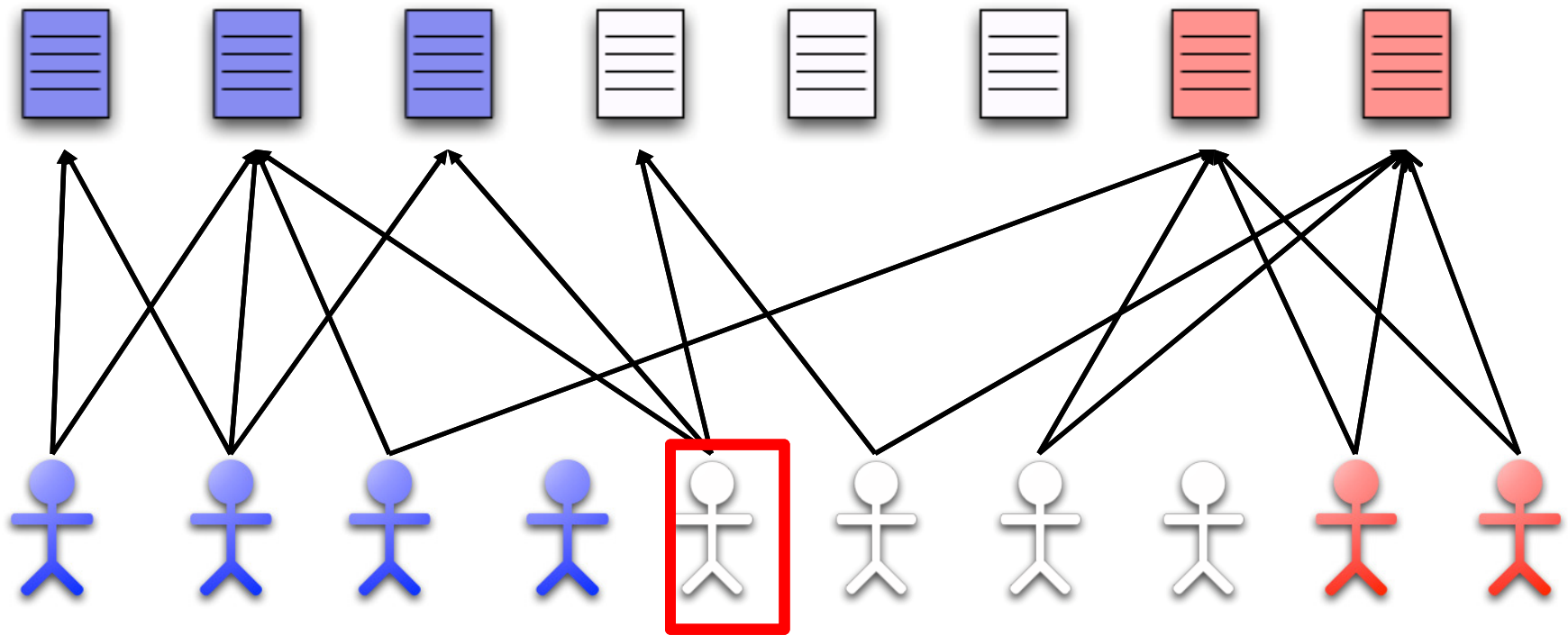
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2011	4134	4266	3883	3083	3879	4055	4386	46				
2007					305	255	2487	2334	2308	2013	2484	2555

Group Email Addresses

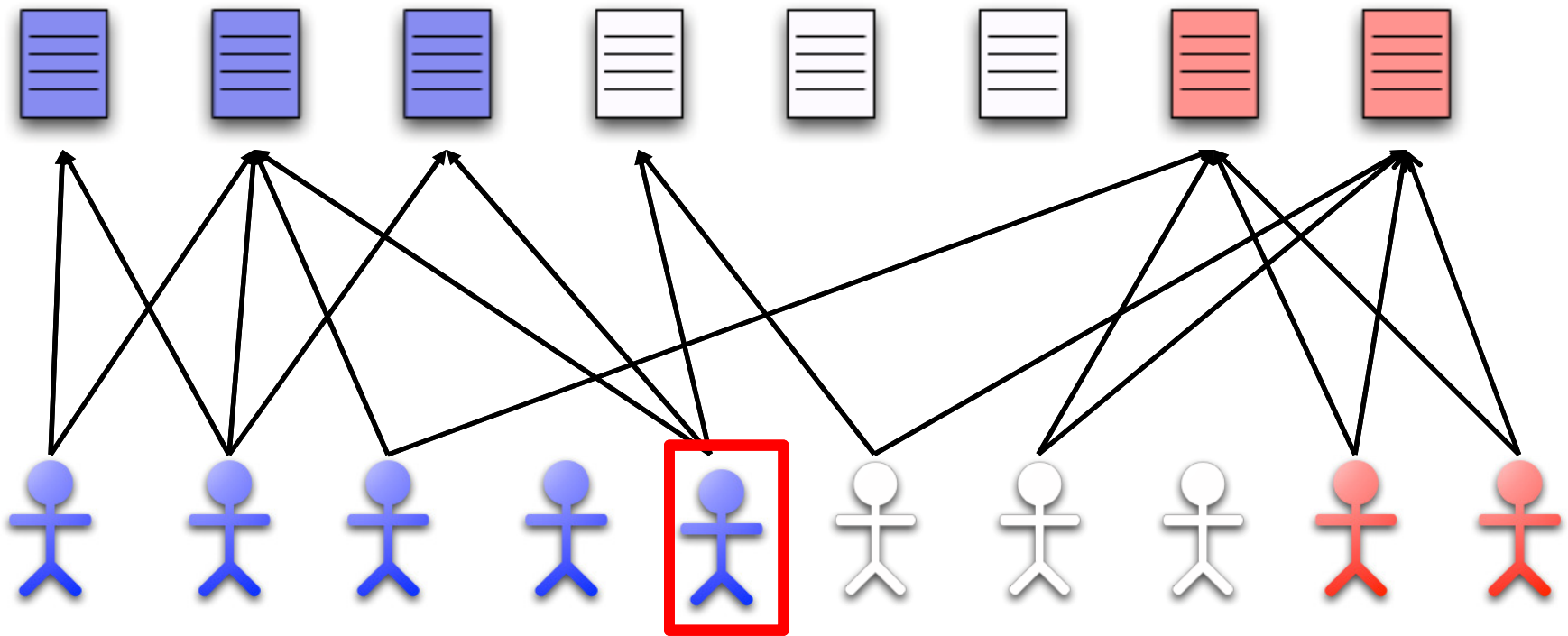
Post message: DiggPatriots@yahoogroups.com
 Subscribe: DiggPatriots-subscribe@yahoogroups.com
 Unsubscribe: DiggPatriots-unsubscribe@yahoogroups.com
 List owner: DiggPatriots-owner@yahoogroups.com

[Join This Group!](#)

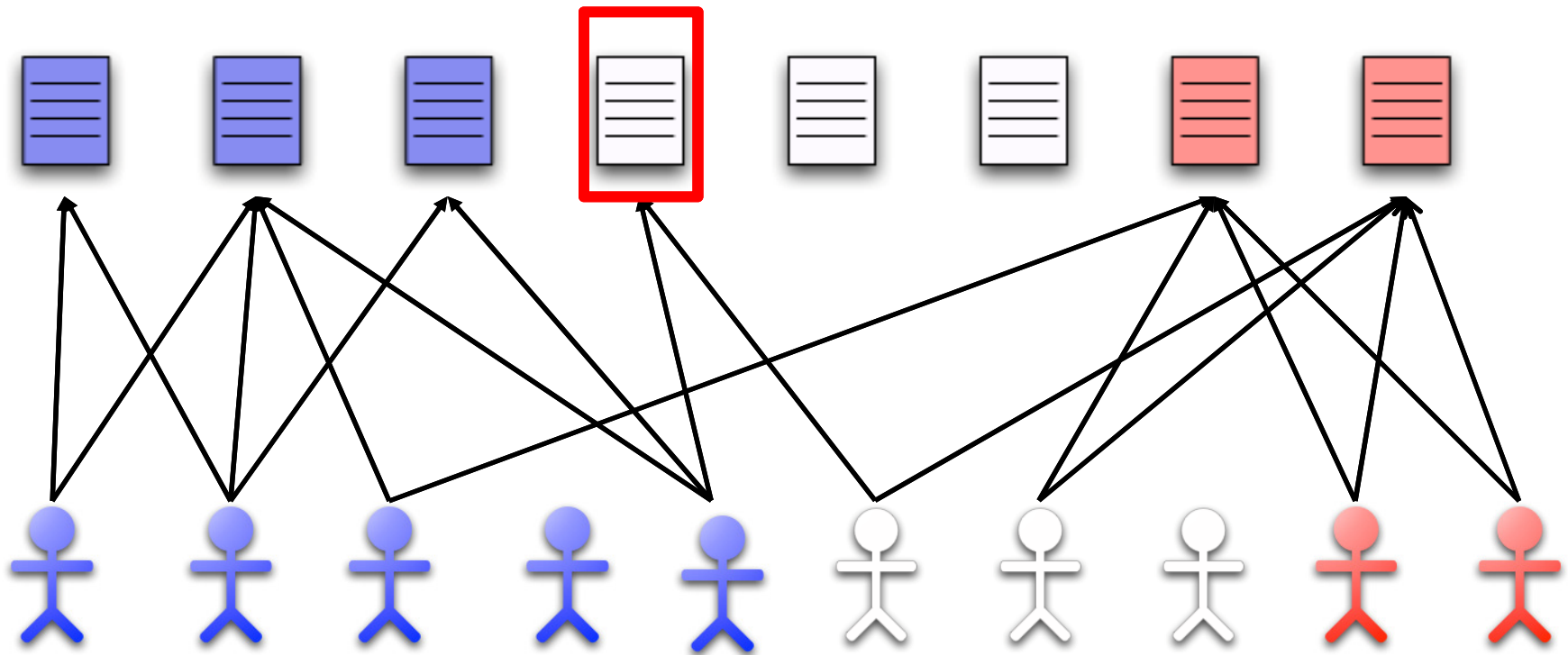
Expanding Coverage By Propagation



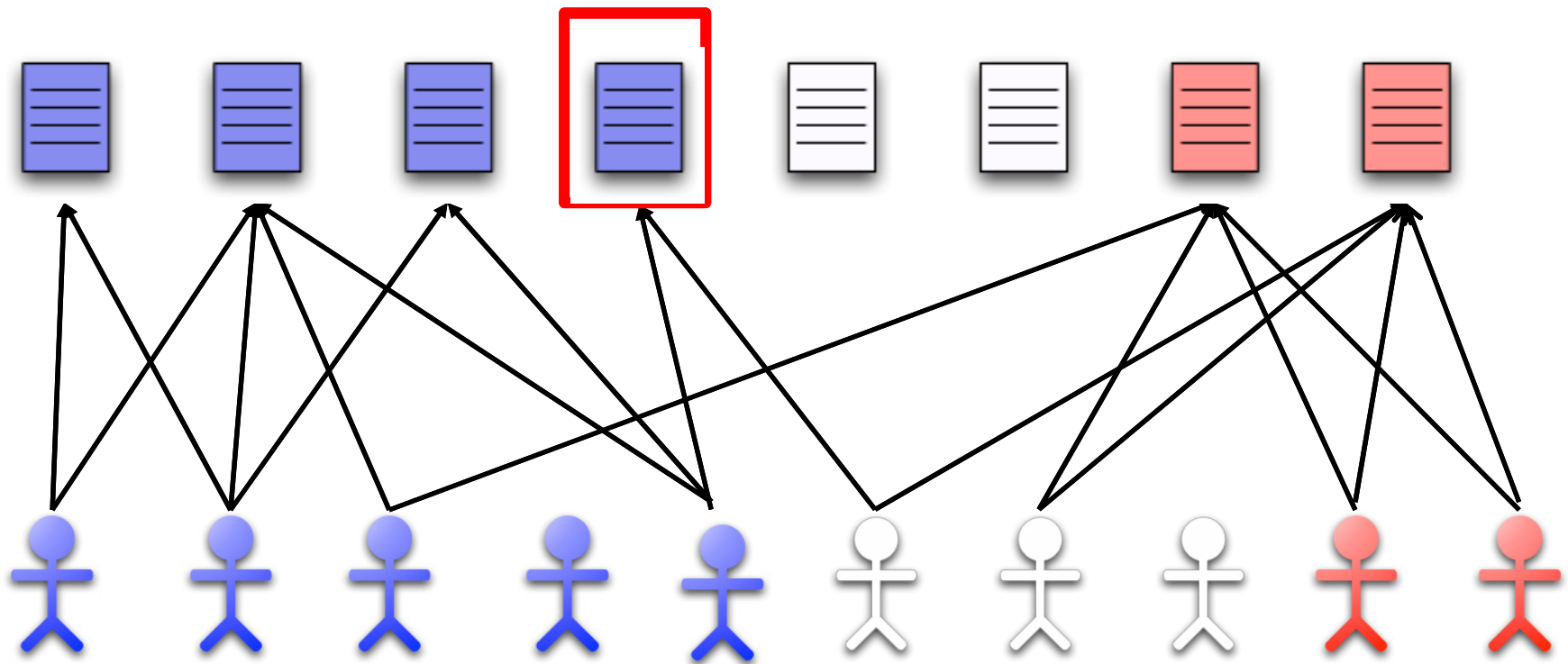
Expanding Coverage By Propagation



Expanding Coverage By Propagation



Expanding Coverage By Propagation



A Semi-Supervised Learning Problem

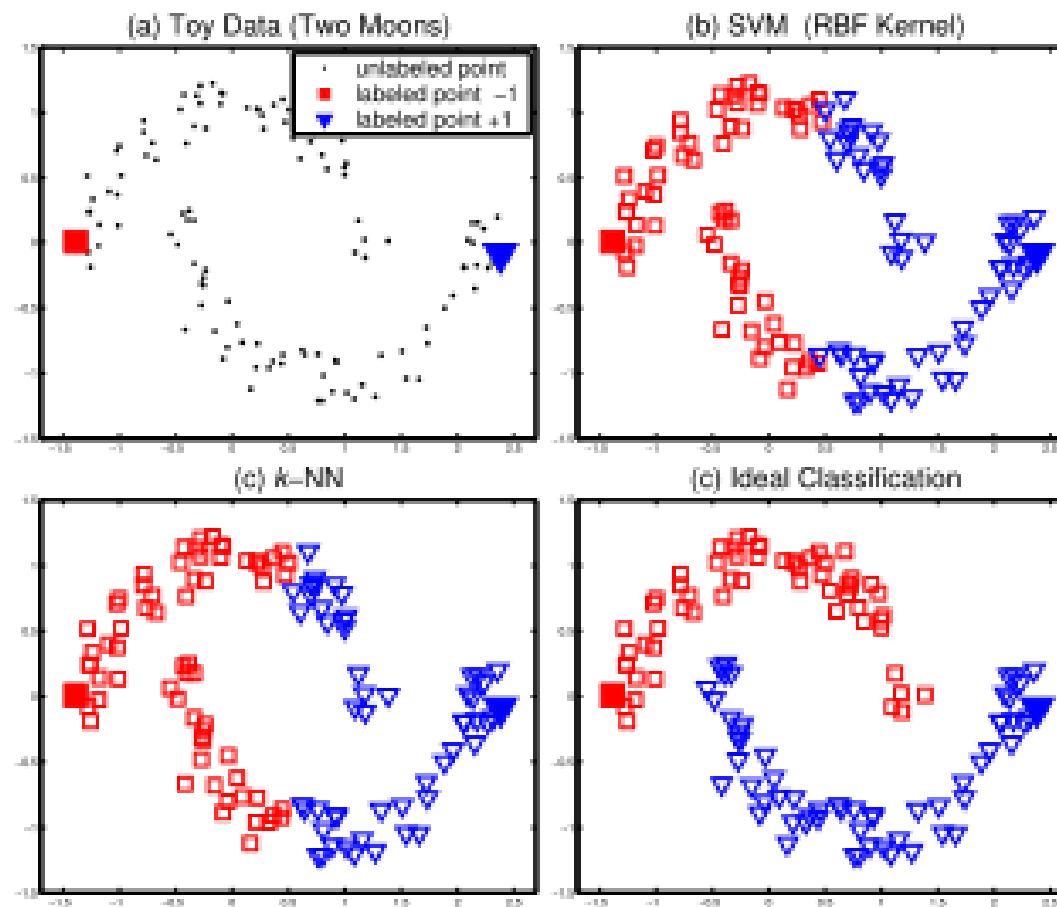
- Some items labeled
- Learn to label the rest
 - Not based on item content features
 - Based on distances/relationships between items

Outline

- Propagation algorithms
- Tuning parameters
- Tuning with additional data sources
- Testing against held out data

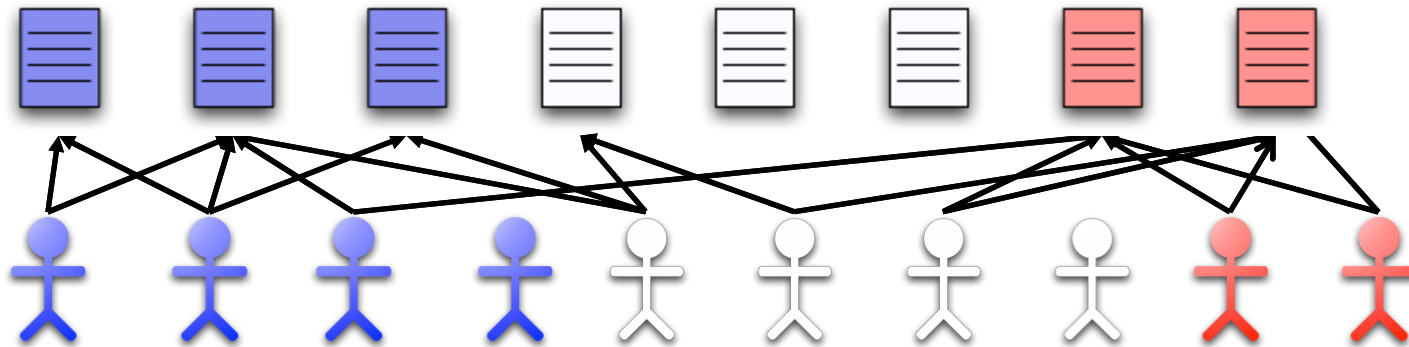
Zhou et al 2004

- "Learning with Global Consistency"



Mapping to Our Case

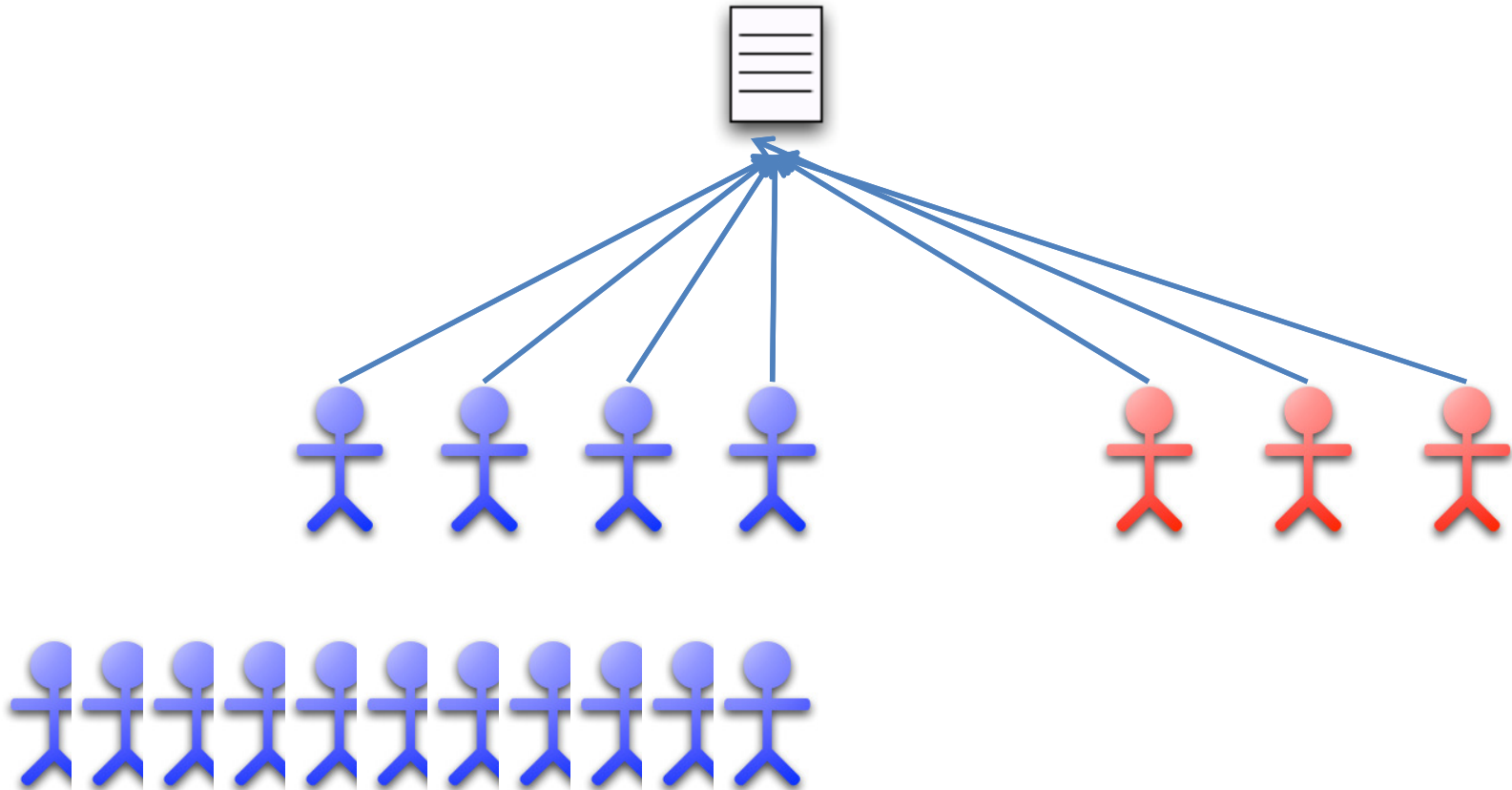
- Choose a continuous (red, blue) score for each node that minimizes a cost function
 - Cost for color distance on adjacent nodes
 - Cost for changing originally labeled nodes
- Equivalently: fixed point of color propagation, with mix of last iteration and original



Desired Behavior (Axiom)

- Color item blue if disproportionately liked by blue people
 - More blues than reds like it
- Color person blue if disproportionately likes blue items
 - Likes more blues than reds

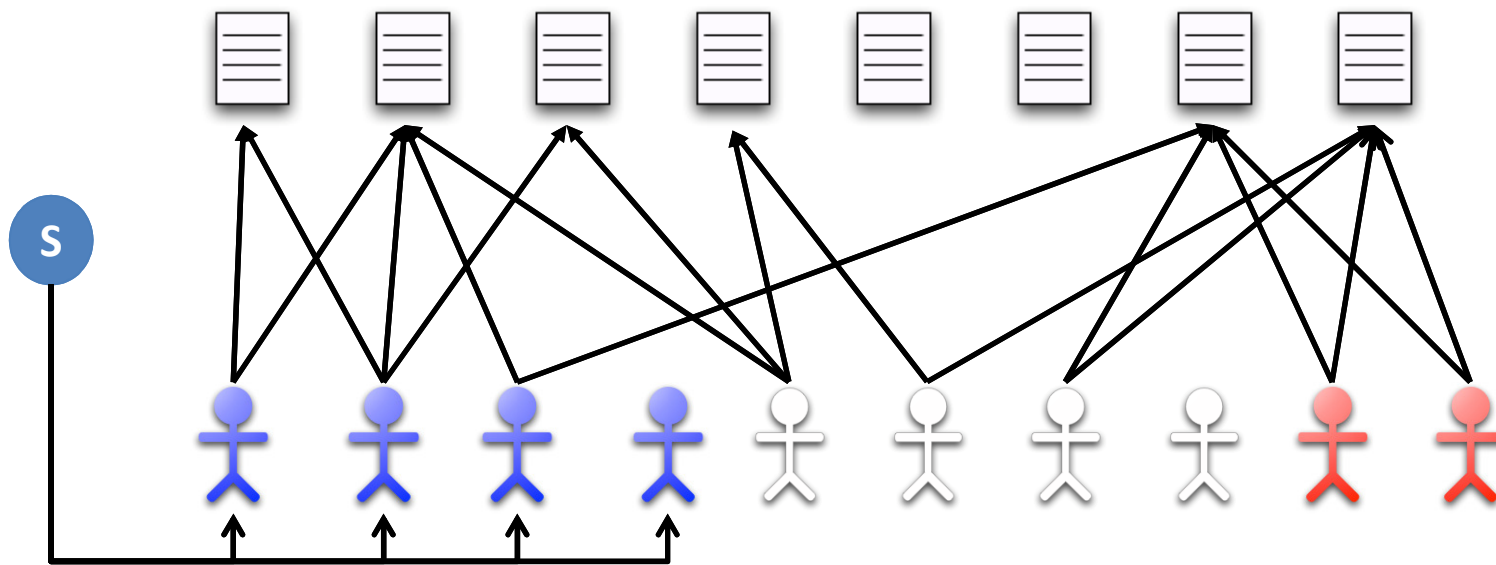
Blue or Red?



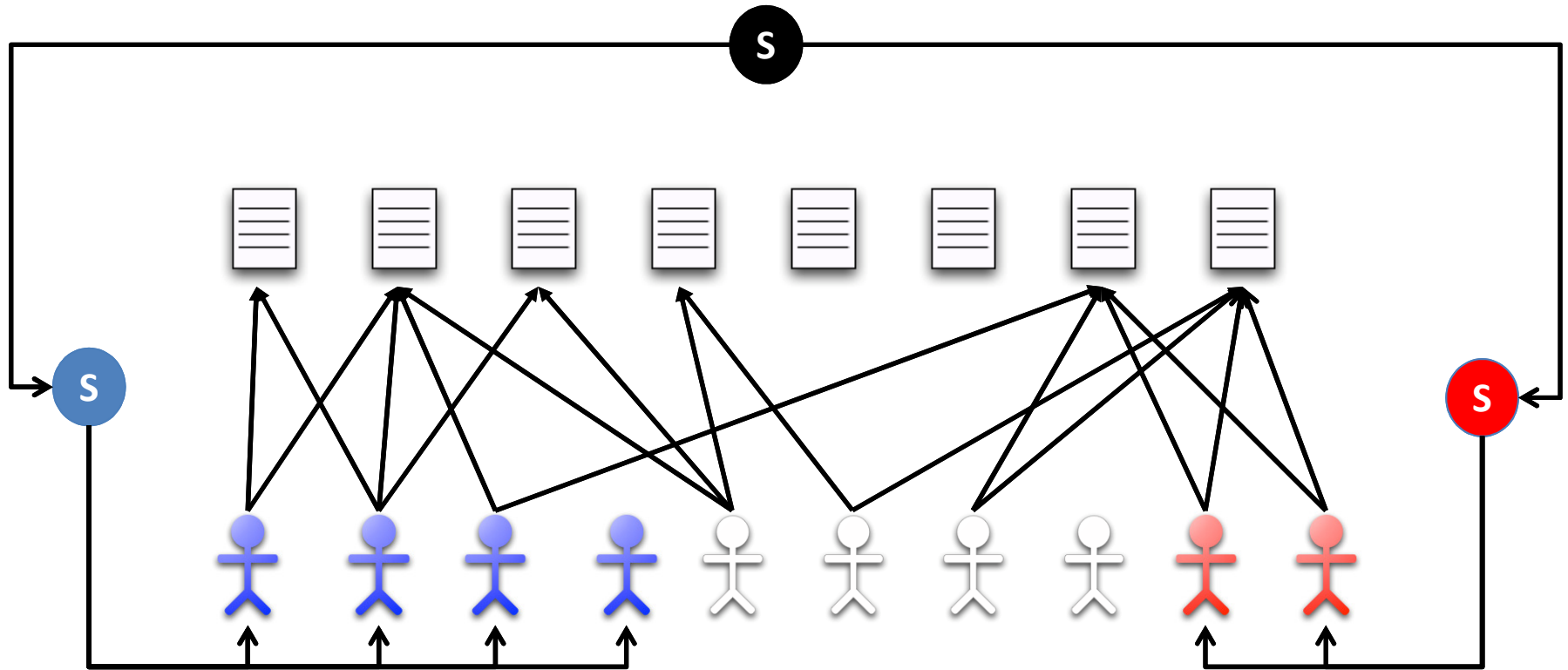
Desired Behavior (Axiom)

- Color item blue if disproportionately liked by blue people
 - ~~More blues than reds like it~~
 - Higher proportion of blues like it than the proportion of reds who like it
- Color person blue if disproportionately likes blue items
 - ~~Likes more blues than reds~~
 - Likes a higher fraction of blue items than of red items

Random Walk With Restart



Random Walk With Restart

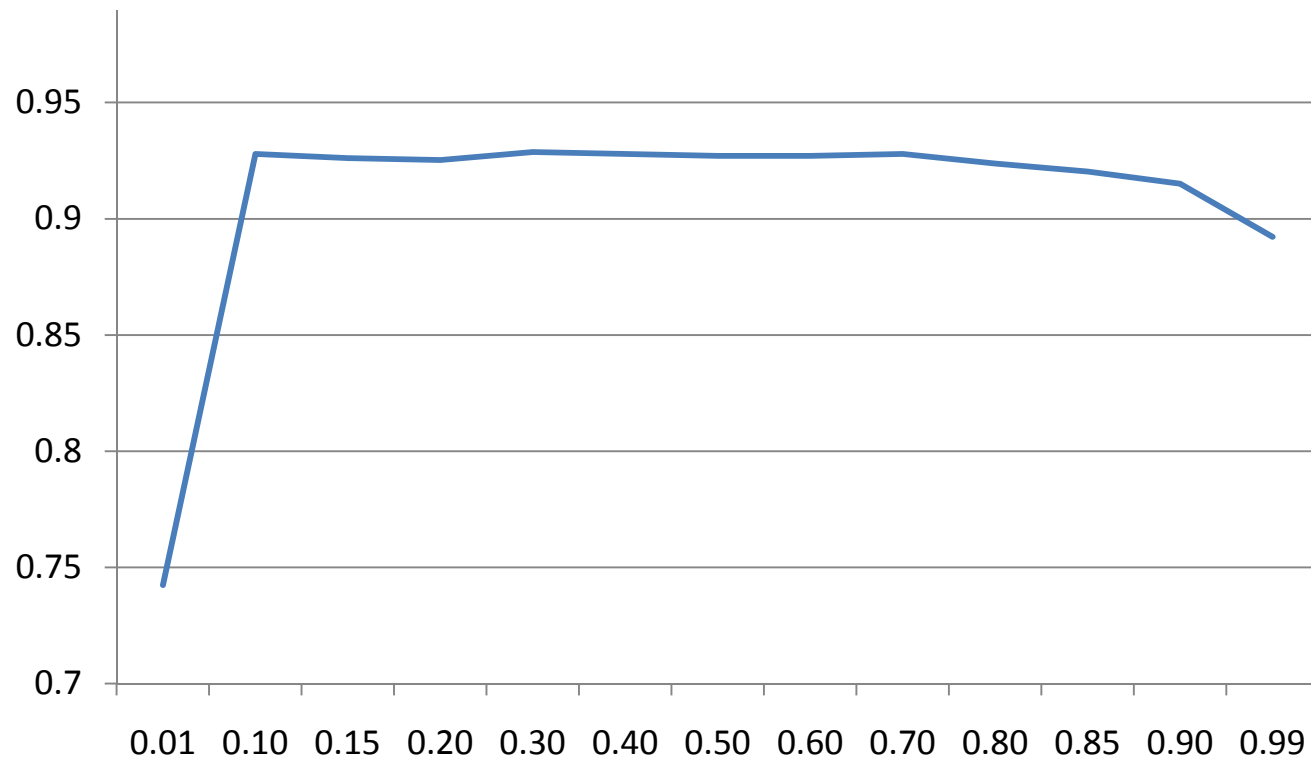


- Classify
 - blue if $\text{Score}_b > \theta \text{Score}_{\text{all}}$
AND NOT $\text{Score}_r > \theta \text{Score}_{\text{all}}$

TUNING PARAMETERS

Tuning α

- Ten-fold cross-validation

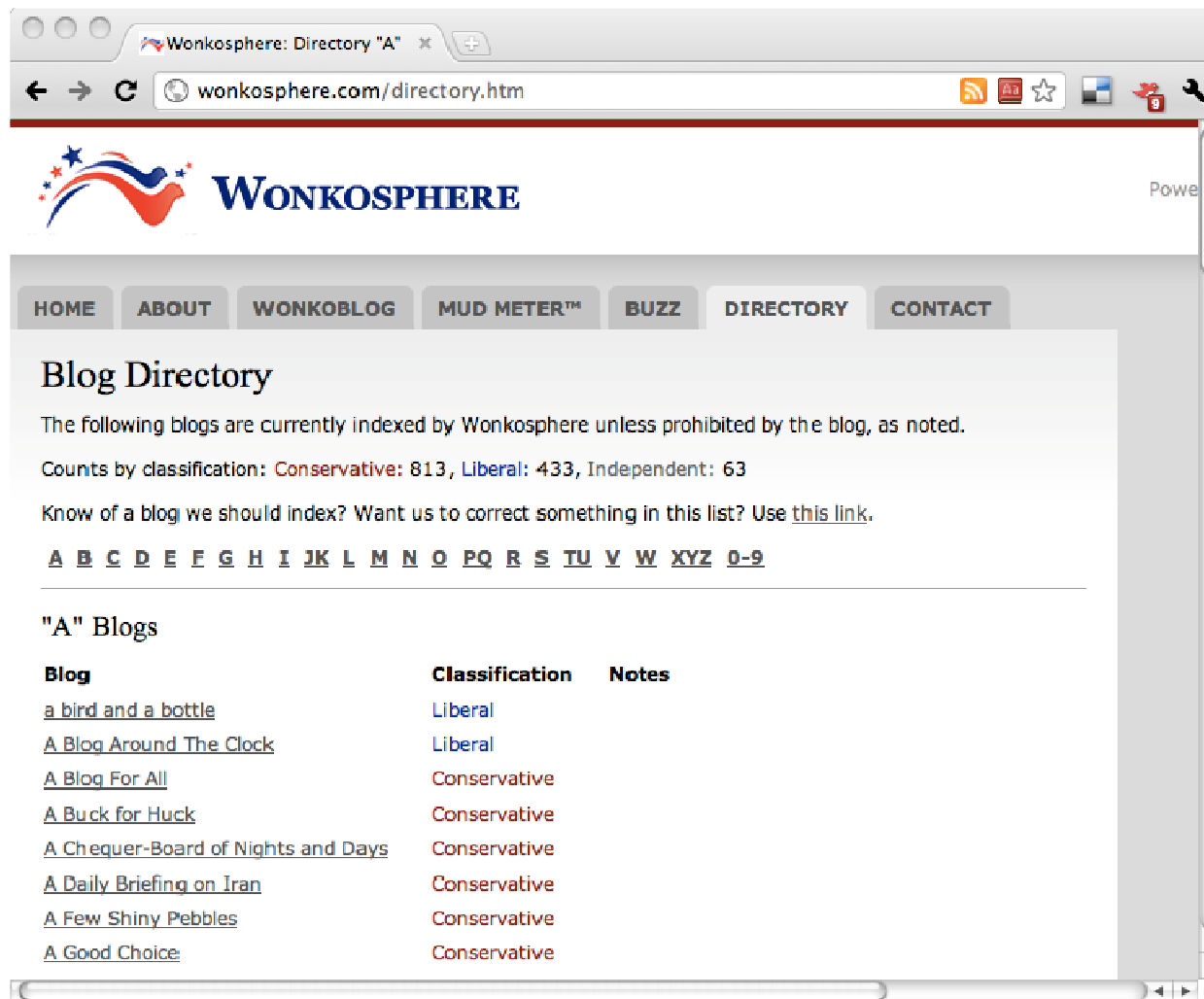


Tuning θ

- Very few labeled nodes as both/neither
 - $\text{Score}_b > \text{Score}_{\text{all}} \iff \text{Score}_r < \text{Score}_{\text{all}}$
- Maximize accuracy at $\theta=1$
 - Everything classified as red XOR blue
- Higher θ yields more "purple" classifications
 - Might be useful if training data had some purples

TUNING WITH ADDITIONAL DATA SOURCES

Externally Classified Blogs

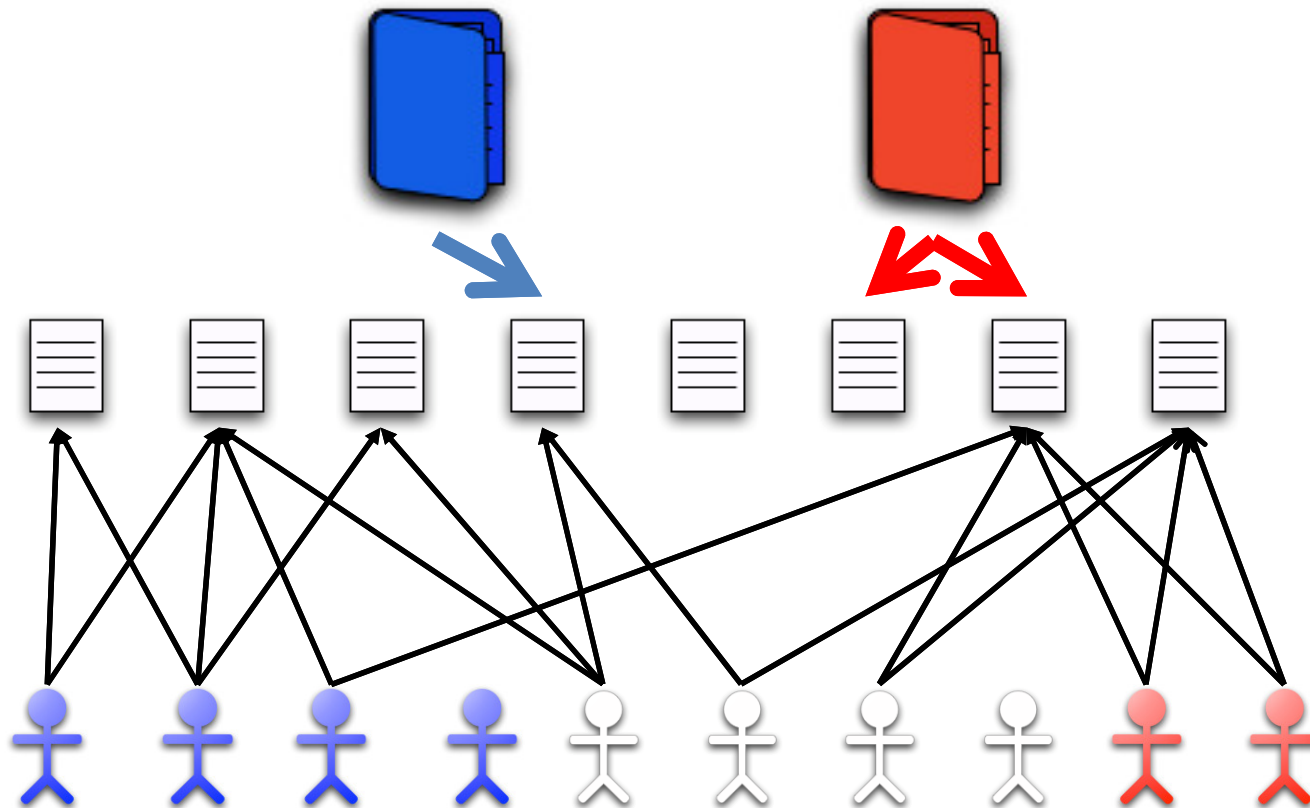


The screenshot shows a web browser window with the address bar displaying "wonkosphere.com/directory.htm". The page features a navigation menu with links: HOME, ABOUT, WONKOBLOG, MUD METER™, BUZZ, DIRECTORY, and CONTACT. The main content area is titled "Blog Directory" and includes a sub-header: "The following blogs are currently indexed by Wonkosphere unless prohibited by the blog, as noted." Below this, it states "Counts by classification: Conservative: 813, Liberal: 433, Independent: 63". A link is provided for users to suggest or correct blog listings: "Know of a blog we should index? Want us to correct something in this list? Use [this link](#)." A horizontal menu lists the alphabet and numbers: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, 0-9. The section "A" Blogs is expanded, showing a table of results.

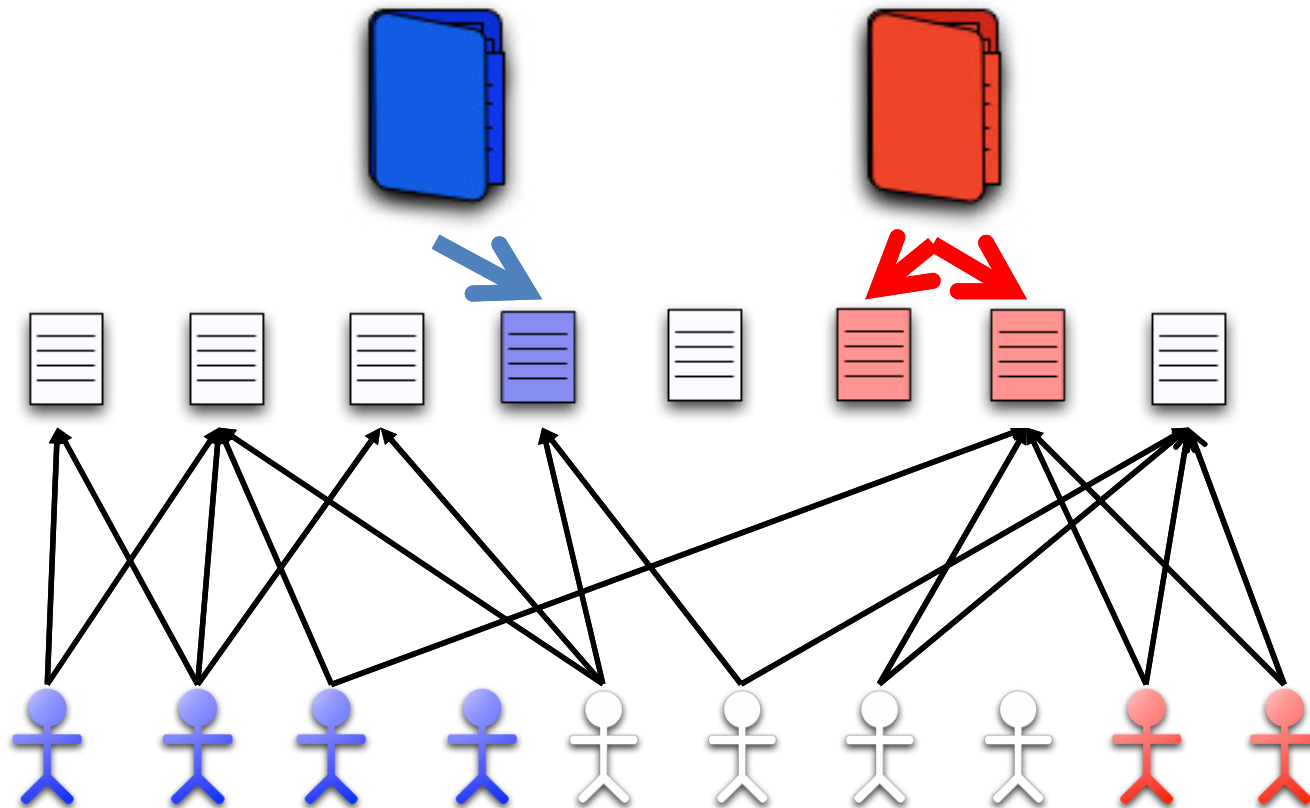
Blog	Classification	Notes
a bird and a bottle	Liberal	
A Blog Around The Clock	Liberal	
A Blog For All	Conservative	
A Buck for Huck	Conservative	
A Chequer-Board of Nights and Days	Conservative	
A Daily Briefing on Iran	Conservative	
A Few Shiny Pebbles	Conservative	
A Good Choice	Conservative	



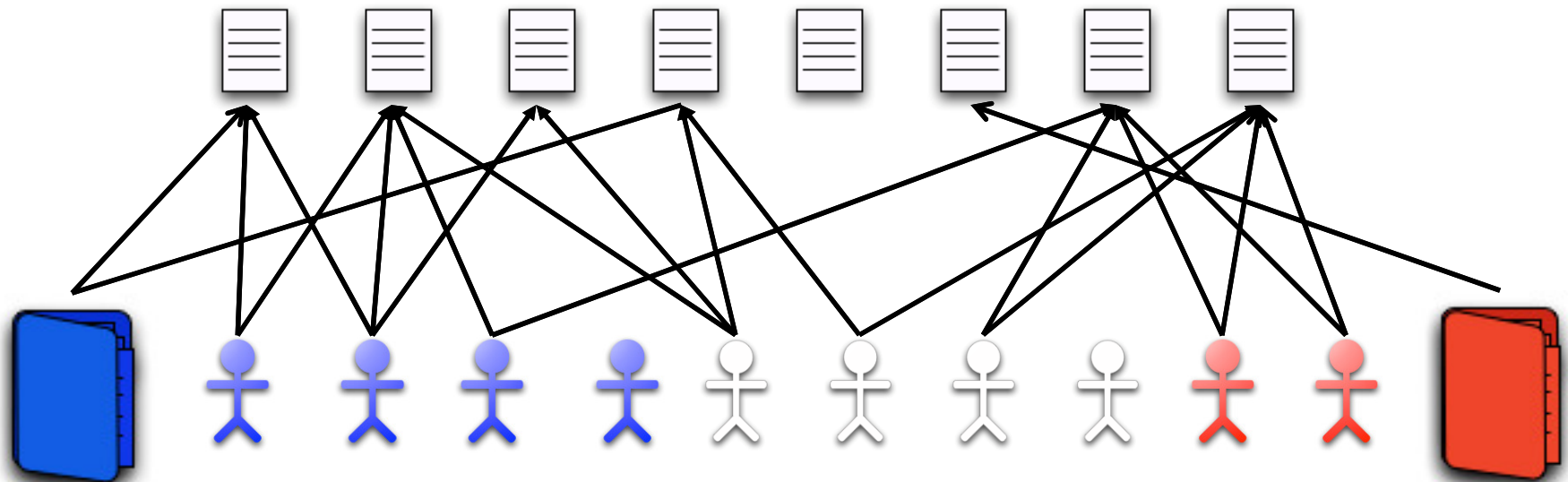
Items In Those Blogs Inherit Labels



Items In Those Blogs Inherit Labels



Blog Links as Votes



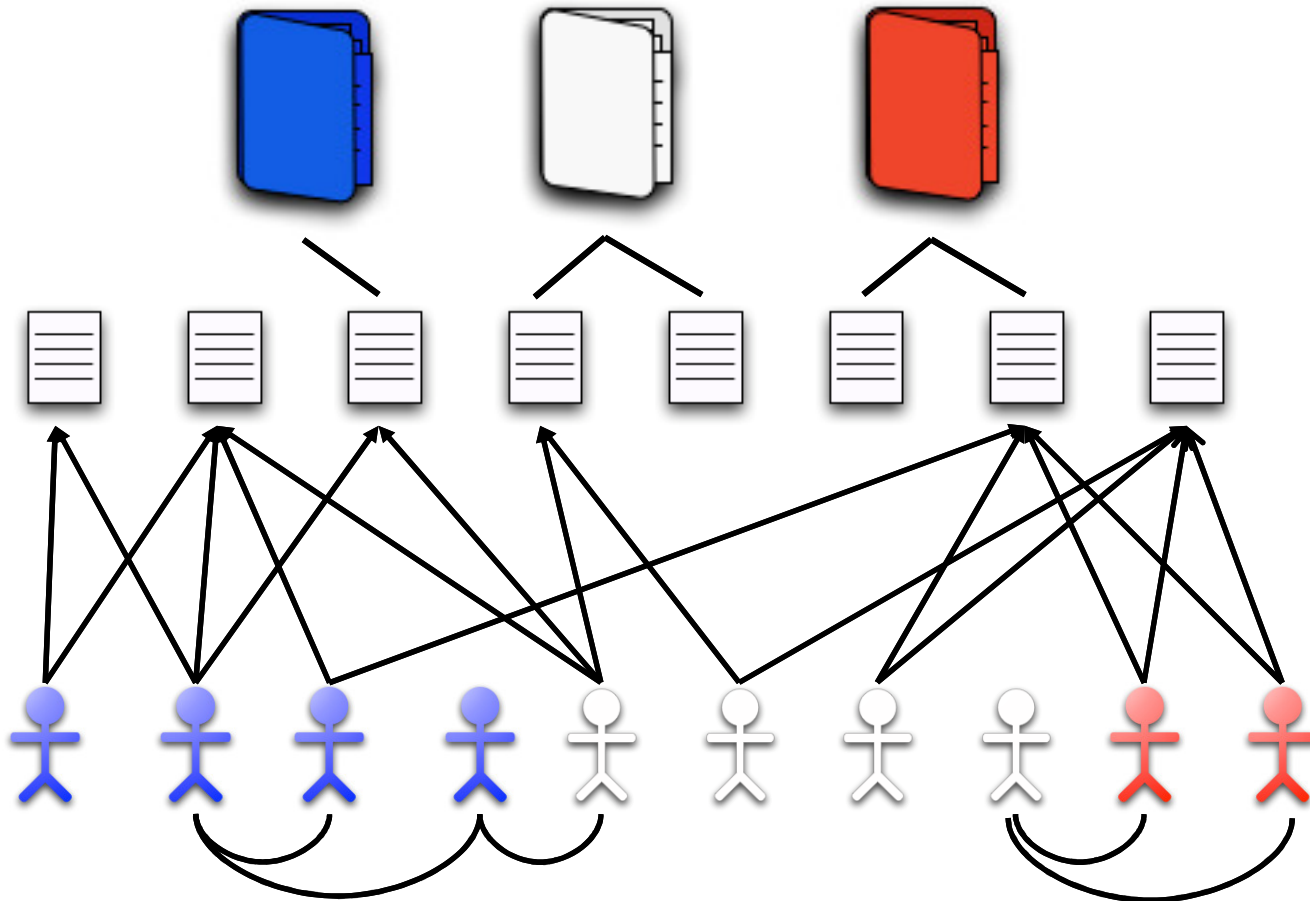
Additional Datasets

- Baseline: only labels for known raters

		add “blog links as votes”	
		No	Yes
add “items from labeled blogs”	No	94.8%	
	Yes	95.4%	92.9%

Additional Domain Data

- "Domain" nodes and "Friend/follower" links

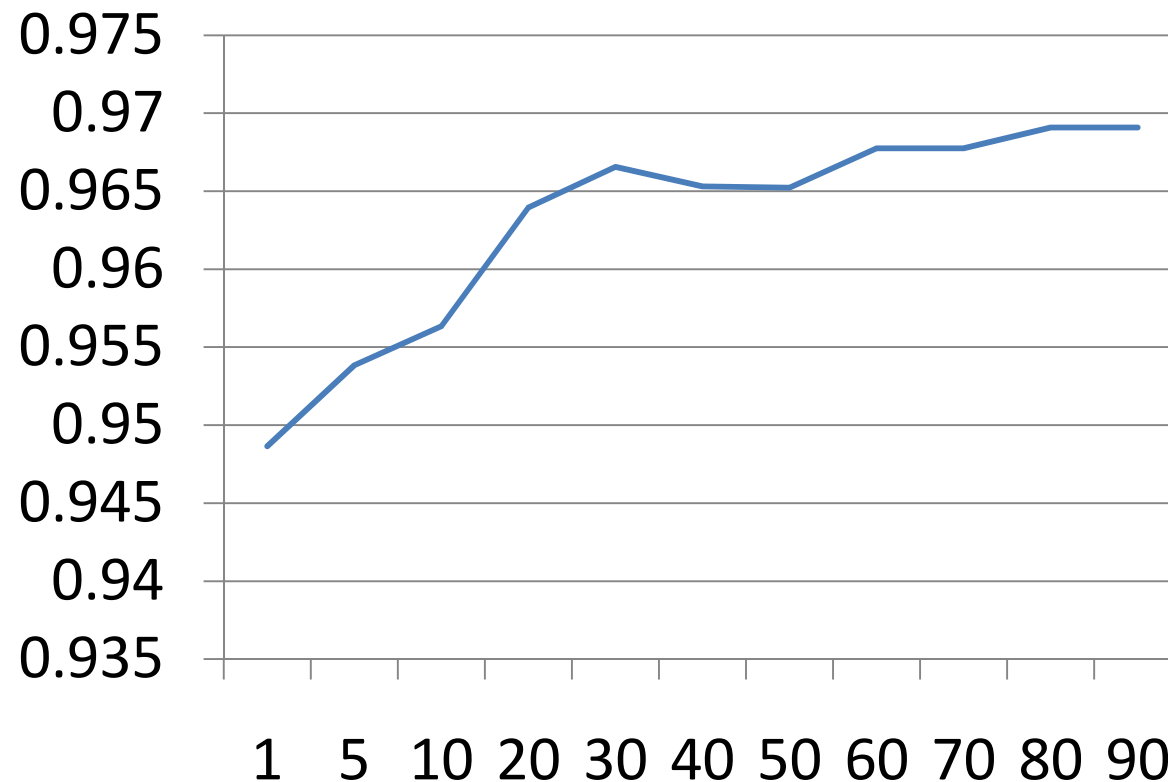


Results

		“domain” links	
		No	Yes
“friendship” links	No	94.9%	94.9%
	Yes	95.1%	95.0%

Weighted Domain Links

- Domain link = n digg votes



TESTING AGAINST HELD OUT DATA

Test Data

- Stories
 - Read and labeled by Turkers
- People
 - User's comments read by research team

'Cadillac Tax' in Health Plan Would Hit Middle Class Hard

7 months ago



Is the article above:

- ☐ **Liberal?** The article needs to have some part of it that satisfies any of the following:
- argues from a liberal perspective, and/or uses liberal ideological arguments (e.g., inequality is bad, social safety net is important, government should not interfere with personal life style choices, etc.)
 - presents facts selectively to promote liberal perspective
 - attacks conservative positions, perspectives, or people from a liberal perspective
- ☐ **Conservative?** The article needs to have some part of it that satisfies any of the following:
- argues from a conservative perspective, and/or uses conservative ideological arguments (e.g., big government is bad, taxes should be low, government should not interfere with markets, etc.)
 - presents facts selectively to promote conservative perspective
 - attacks liberal positions, perspectives, or people from a conservative perspective
- ☐ **Both?** The article has a part that satisfies the criteria for "liberal" and a part that satisfies the criteria for "conservative"
- ☐ **Neither (non-partisan)?** The article needs to satisfy any of the following:
- does not satisfy criteria of "liberal" or "conservative"
 - presents simple facts with no hint of political opinions
- ☐ **Broken link?** The article needs to satisfy any of the following:
- links to nothing
 - links to something other than what its title suggested

Want to work on this HIT?

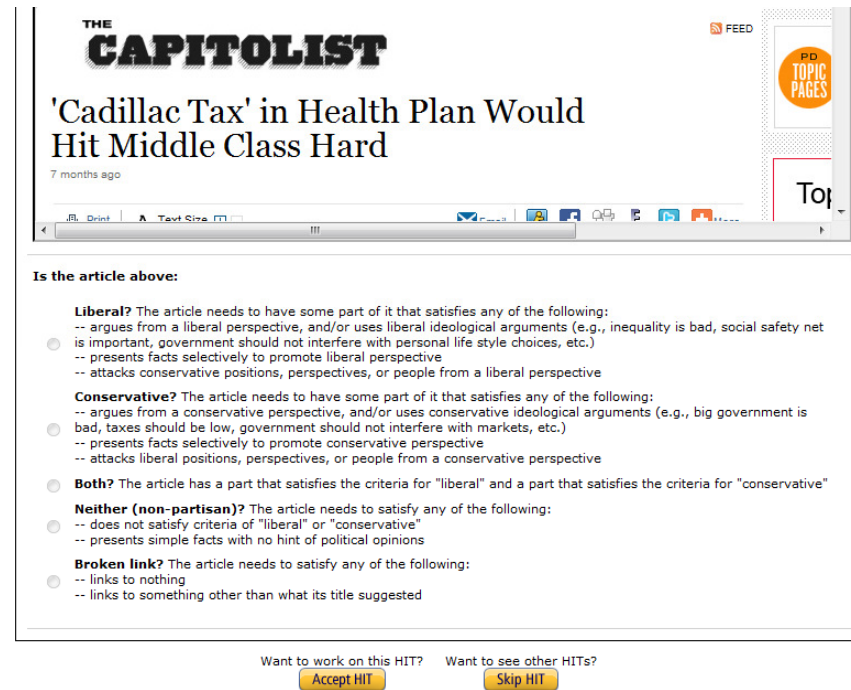
Accept HIT

Want to see other HITs?

Skip HIT

Stories Labeled by Mechanical Turkers

- 1000 stories
- Qualification test to hire turkers
- 3 self-identified liberal and 3 self-identified conservative for each story
- Only used 307 labeled stories with 6/6 agreement that are either red or blue (73 red, 234 blue)



The screenshot shows a HIT interface for labeling a news article. At the top, the article is from 'THE CAPITOLIST' and is titled ''Cadillac Tax' in Health Plan Would Hit Middle Class Hard'. Below the title, it says '7 months ago'. The interface includes a browser window showing the article and a list of criteria for labeling the article as Liberal, Conservative, Both, Neither, or Broken link. At the bottom, there are two buttons: 'Accept HIT' and 'Skip HIT'.

THE CAPITOLIST

'Cadillac Tax' in Health Plan Would Hit Middle Class Hard

7 months ago

Is the article above:

- ☐ **Liberal?** The article needs to have some part of it that satisfies any of the following:
 - argues from a liberal perspective, and/or uses liberal ideological arguments (e.g., inequality is bad, social safety net is important, government should not interfere with personal life style choices, etc.)
 - presents facts selectively to promote liberal perspective
 - attacks conservative positions, perspectives, or people from a liberal perspective
- ☐ **Conservative?** The article needs to have some part of it that satisfies any of the following:
 - argues from a conservative perspective, and/or uses conservative ideological arguments (e.g., big government is bad, taxes should be low, government should not interfere with markets, etc.)
 - presents facts selectively to promote conservative perspective
 - attacks liberal positions, perspectives, or people from a conservative perspective
- ☐ **Both?** The article has a part that satisfies the criteria for "liberal" and a part that satisfies the criteria for "conservative"
- ☐ **Neither (non-partisan)?** The article needs to satisfy any of the following:
 - does not satisfy criteria of "liberal" or "conservative"
 - presents simple facts with no hint of political opinions
- ☐ **Broken link?** The article needs to satisfy any of the following:
 - links to nothing
 - links to something other than what its title suggested

Want to work on this HIT? [Accept HIT](#) Want to see other HITs? [Skip HIT](#)

Labeling Users

- 220 Users with
 - >15 comments;
 - >5 diggs
 - >½ of recent comments on political stories
- 2 undergrad coders
 - One code for each user
 - Code as red (conservative), blue (liberal), gray (unclear or signs of both)
- Kappa score: 0.69
 - .89 excluding users coded by either as gray
 - 6 pre-labeled users included for quality control
 - all coded correctly
- 131 users where coders agreed on pure red or blue
 - (69 red, 62 blue)

Accuracy

Cross-validation on training data	94.90%
On test data	94.5% overall, 92.8% for stories, 98.5% for users

Conclusion

- Can propagate political leanings from a few labeled items and people
- Much room for tuning
 - Algorithm: definition of "disproportionately liked"
 - Data to include
 - ✓ Diggs as votes
 - ✓ Source blog → item label
 - x Blog references as votes
 - x Friend links
 - ✓ Source links

