

The Power of Public Data

Matt Smith

<http://m.smithworx.com>

Web Startup Group

April 2010



Information and Attention

- ✦ “Internet data heads for 500bn gigabytes” (May 2009)
- ✦ Transparency and Open Government
- ✦ Social media data is rampant

<http://www.guardian.co.uk/business/2009/may/18/digital-content-expansion>

http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/

Data to Knowledge

✦ Data



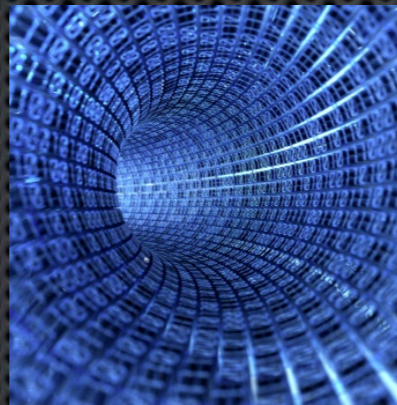
✦ Information



✦ Knowledge

10480	15011	01536	02011	81347	91645
22368	46573	25595	85393	30985	89198
24130	48360	22527	87265	76363	64509
42167	93093	06243	61680	07365	16375
37570	30975	81937	16686	06121	91782
77921	06997	11008	42751	27755	53498
39552	72905	56420	69694	98372	31076
96301	91977	05463	07972	18375	20522
89579	14342	63661	10281	17453	18103
85475	36857	53342	53988	53360	59533
28018	60678	86231	33276	70267	79038
63553	40961	48235	03427	49326	69445
09429	93969	52636	92737	88974	33488
10365	61129	87529	85689	48237	52267
07119	97336	71048	08178	77233	13916
51065	12765	51821	51259	77452	16308
02368	21382	52404	60268	89368	19885
01011	54092	33362	94904	31273	04146
52162	53916	46369	58586	23216	14513
07056	97628	33767	09698	42688	06591
48663	91245	85828	14346	09172	30168
54164	58492	22421	74103	47070	25306
32639	32363	05567	24200	13363	38005
29334	27001	87637	87308	58731	00266
02488	53062	28834	07351	19731	92420

Population of Utah

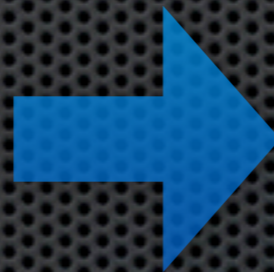
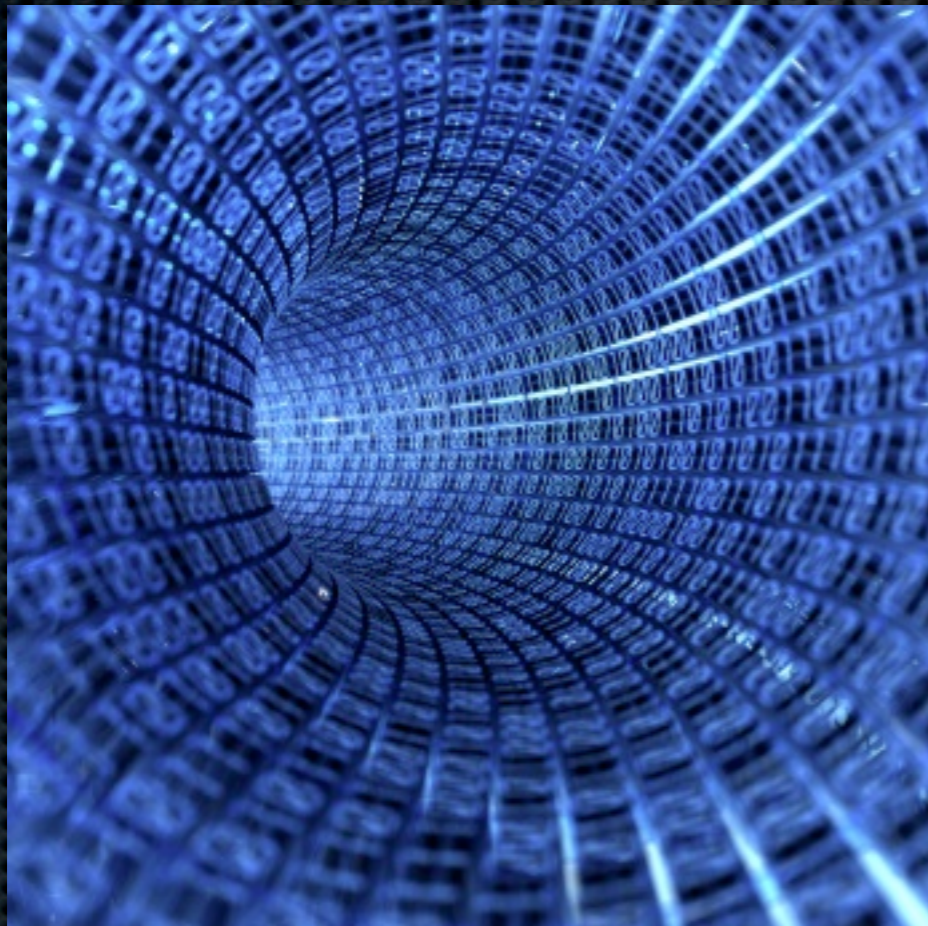


Socio-demographics of Utah



How to do business in Utah

Data Mining for Knowledge



Meaningless

Digestible and Useful

Data Mining Tasks



- ✦ **Classification** - model data to predict future instances
- ✦ **Clustering** - organize information into natural groups
- ✦ **Regression** - find optimal function that fits the data
- ✦ **Association Rules** - identify relationships within data
- ✦ **Link Mining** - consider links among entities for above

Government - Federal



Agriculture
Arts, Recreation, and Travel
Banking, Finance, and Insurance
Births, Deaths, Marriages, and Divorces
Business Enterprise
Construction and Housing
Education
Elections
Energy and Utilities
Federal Govt. Finances & Employment
Foreign Commerce and Aid
Geography and Environment
Health and Nutrition
Income, Expend., Poverty, and Wealth
Information and Communications

International Statistics
Labor Force, Employment, and Earnings
Law Enforcement, Courts, & Prisons
National Security and Veterans Affairs
Natural Resources
Other
Population
Prices
Science and Technology
Social Insurance and Human Services
State Govt. Finances & Employment
Transportation
Wholesale and Retail Trade

Political Contributions

As featured in the [Google Maps Election Gallery](#) [Send](#) [Link to this page](#)

ATTORNEY SELF-EMPLOYED	Mitt Romney \$2,300	DRIVE Provo UT
H. Ross INFORMATION REQUESTED PER BEST E INFORMATION REQUESTED PER BEST E	Updated Q4/2007 Mitt Romney \$2,300	3737 N 970 E Provo UT
Paul Allen ENTREPRENEUR WORLD VITAL RECORDS	NEW Updated Q1/2008 Mitt Romney \$2,300	1806 N 500 E Provo UT
Kristin Kelly Mudock BUSINESS SELF-EMPLOYED	Updated Q3/2007 Mitt Romney \$2,300	4048 DEVONSHIRE DRIVE Provo UT
Mitch Huhem ENTREPRENEUR SELF EMPLOYED	Updated Q1/2007 Mitt Romney \$2,300	2230 N UNIV PKWY Provo UT
Cindy Nemrow HOMEM.	Updated	
Scott Ber RETIRED		
William Freedman BUSINESS DOWNEAST OUTFITTER		
Susan Pace NONE NONE		
Minnie Mccarney Updated		Provo UT

The Federal Election Commission (FEC) requires that all campaign contributions over \$200 (per donor) be reported publicly. The reported information includes the donor's name, job title, zip code, and even address. All of it, since 2001 is available electronically via FTP at <ftp://ftp.fec.gov/FEC/electronic/>.

EXAMPLE
Download campaign contributions for May:
`wget -N ftp://ftp.fec.gov/FEC/electronic/201003*.zip`

Crime

CrimeSpotting.org

San Francisco
crimespotting

Report data last published by San Francisco Police: 21 days ago

Home Map Crime Reports API About Feedback



TIME OF DAY

Show All | Hide All

Light | Dark [nearest hour]

Commute | Nightlife

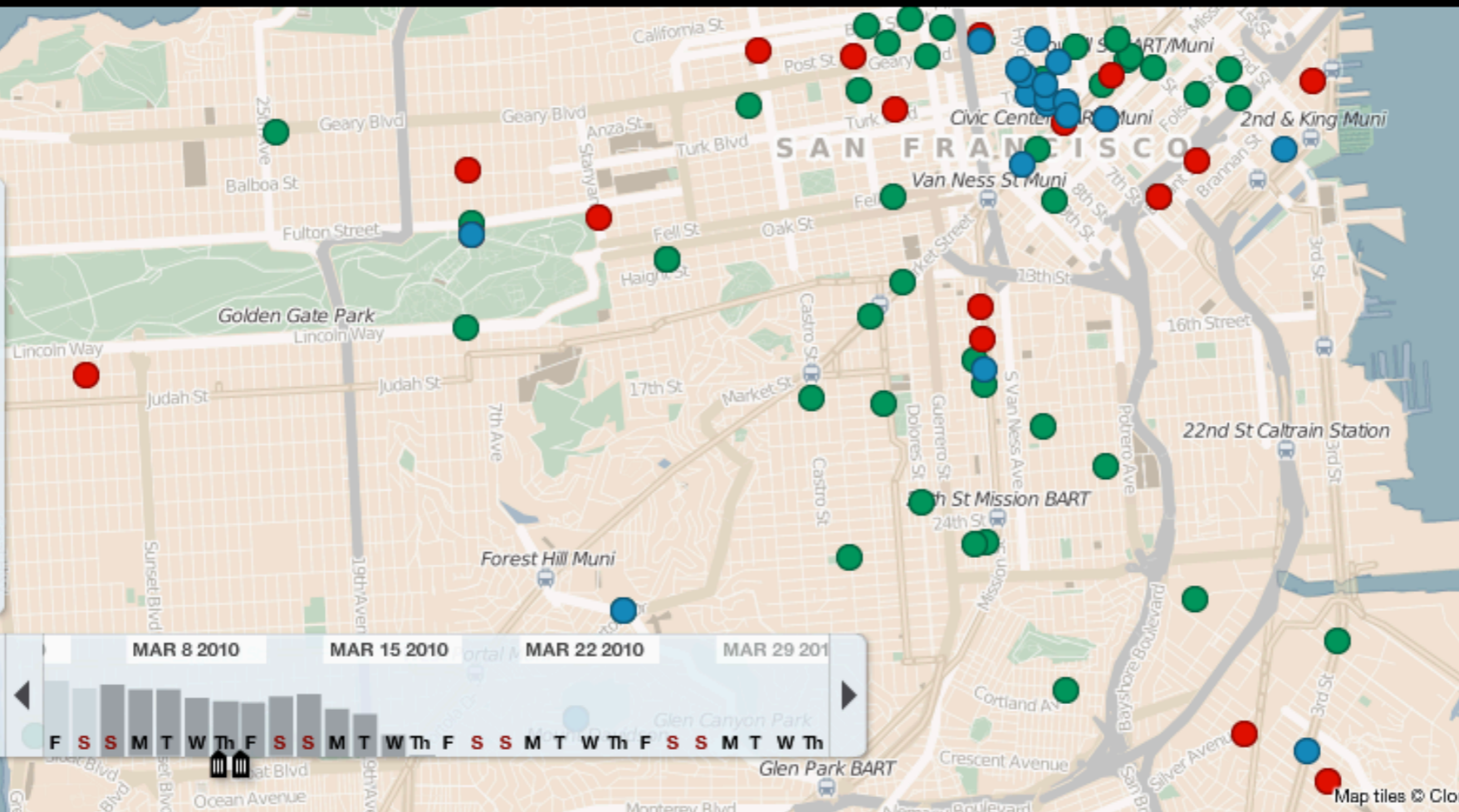
Day | Night | Swing Shift



DATE

Past Week

Mar 2010



CRIME TYPE

Show All | Hide All

- AA Aggravated Assault
- Mu Murder
- Ro Robbery
- SA Simple Assault
- DP Disturbing the Peace
- Na Narcotics
- Al Alcohol
- Pr Prostitution
- Th Theft
- VT Vehicle Theft
- Va Vandalism
- Bu Burglary
- Ar Arson

Map tiles © CloudMade. Map data CC-BY-SA OpenStreetMap.org

Government - State & Local

<http://www.data.gov/statedatasites>

- ✦ States (6):

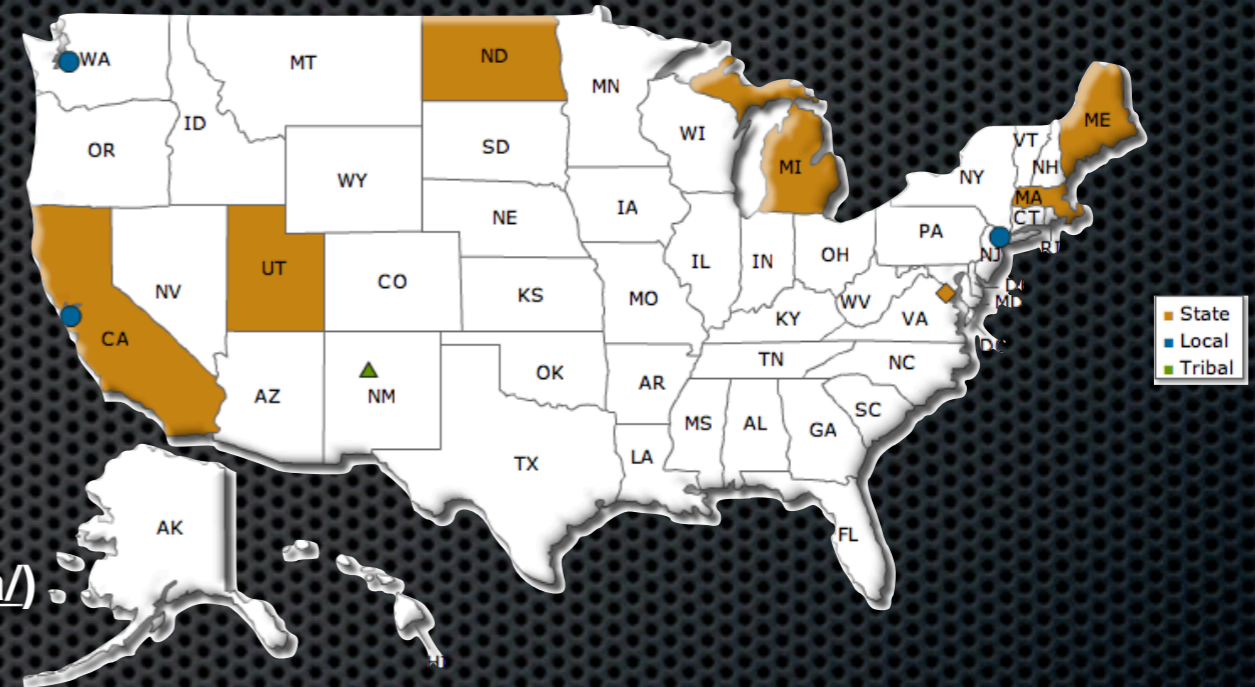
- ✦ Utah (<http://www.utah.gov/data/>)

- ✦ California (<http://www.ca.gov/data/>)

- ✦ North Dakota, Michigan, Maine, Massachusetts

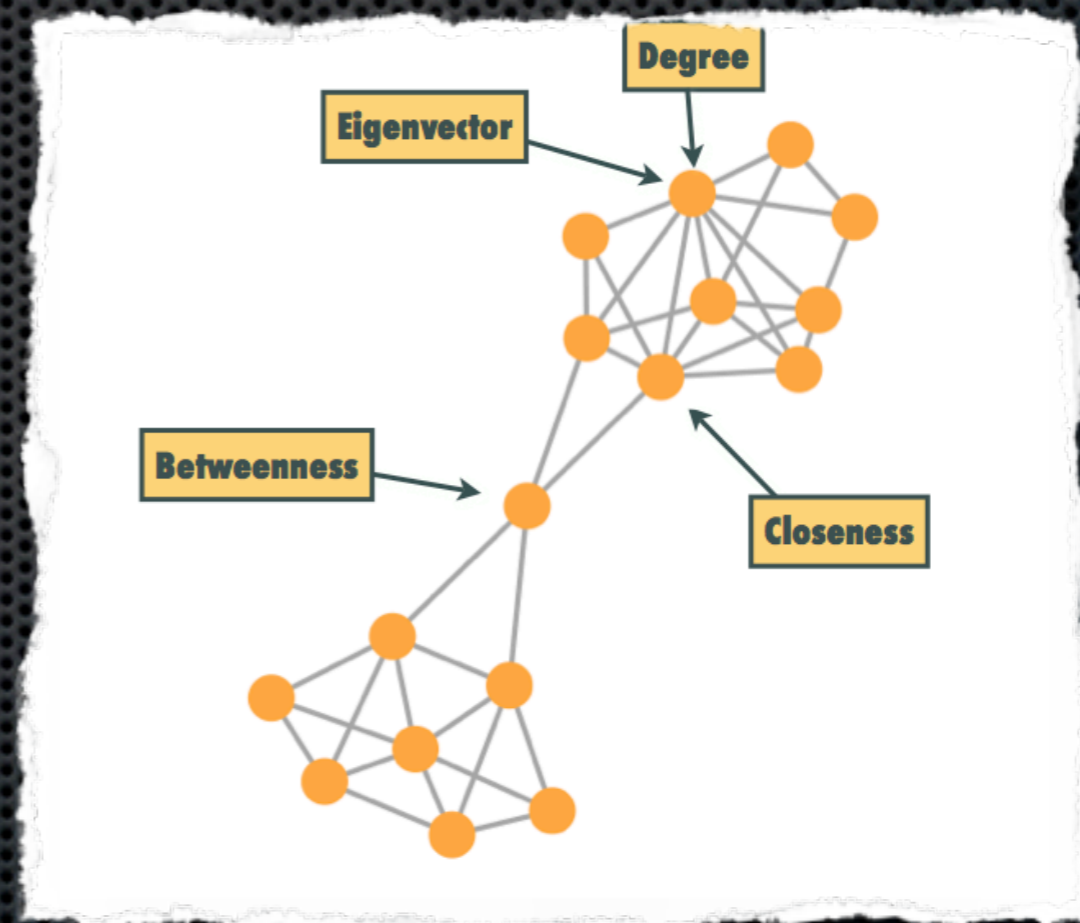
- ✦ Cities:

- ✦ Seattle, New York, San Francisco

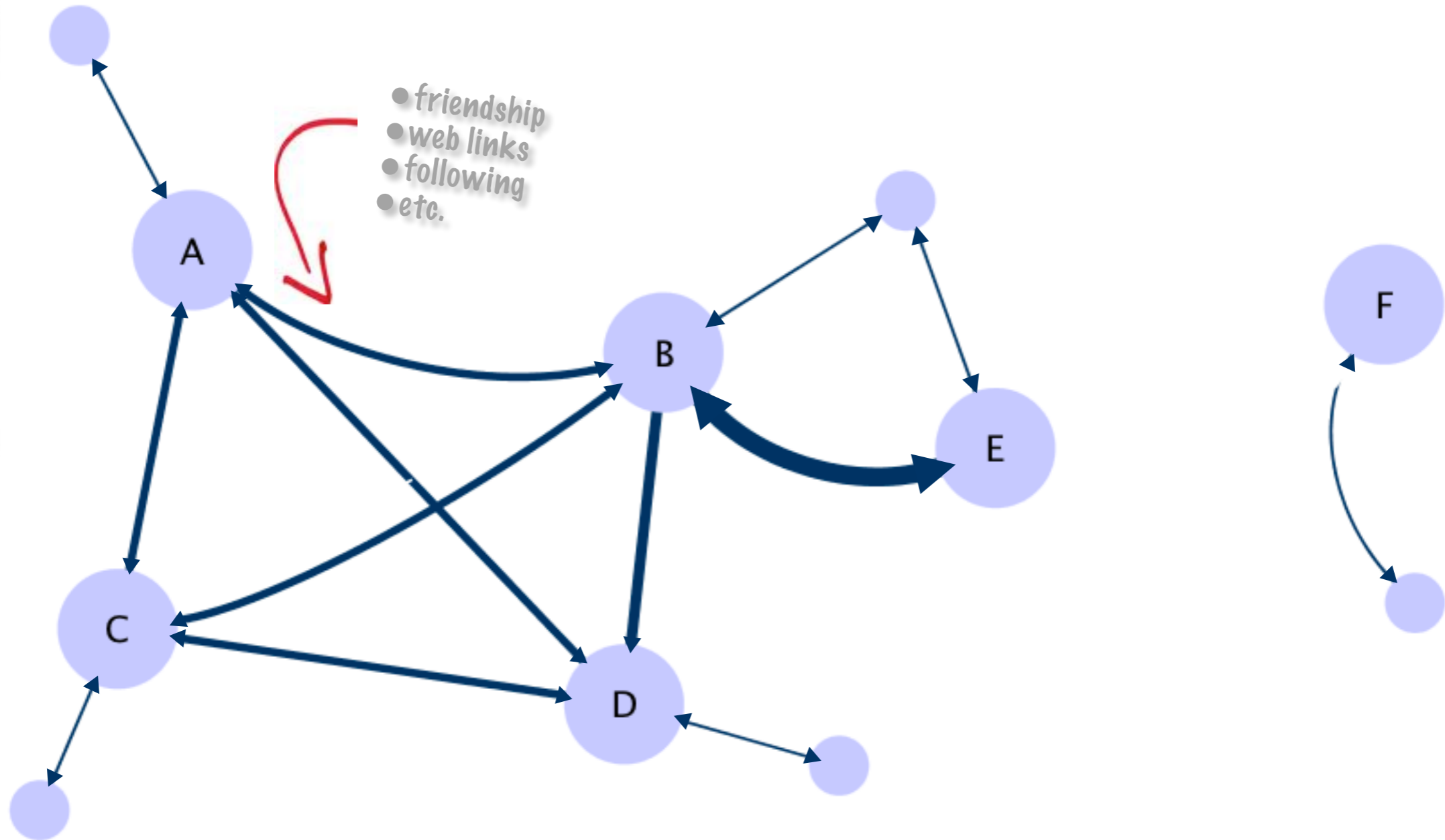


Social Data

- ✦ Twitter
- ✦ Facebook
- ✦ Blogosphere
- ✦ YouTube
- ✦ Flickr



Social Analytics



Other Great Data Sources

- ✦ Wikipedia (<http://download.wikimedia.org/>)
- ✦ IMDB (<http://www.imdb.com/interfaces>)
- ✦ Scriptures (<http://scriptures.nephi.org>)
- ✦ FreeBase (<http://freebase.com>)
- ✦ Flickr (<http://flickr.com>)
- ✦ ScrapMaker Word Lists (<http://scrapmaker.com>)

Conclusions

- ✦ Lots of great data that ripe for use
- ✦ Leverage available data in your next venture
- ✦ Solve government problems
- ✦ Create social tools

Questions & Comments

Ask me now:



Email or Call me:

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

Connect:

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

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

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

Twitter: [@smithworx](https://twitter.com/smithworx)