



ILS INSTITUTE

UNIVERSITY AT ALBANY State University of New York

INTRODUCTION TO SOCIAL COMPUTING

Tomek Strzalkowski and Samira Shaikh

1

SOCIAL COMPUTING

- A research area on the intersection of computer and social science
- Deals with processing and understanding of human generated data
 - Language
 - Behavior
 - Opinions
 - Preferences
 - Sentiment
 - Beliefs
 - Relationships
- Reveal social phenomena in human interactions

THE NAME GAME

- Social computing vs. computational social science?
 - Social computing =? Theory-free big data exploitation of social data
 - E.g., People with longer names more likely to be neurotic
 - Can't separate real from accidental
 - Computational social science = theory based exploitation of big data
 - Statistical validation of all claims, must meet significance criteria
 - All results verified against reliable human judgment
- Social Computing is more catchy 😊
 - We will always assume this to mean CSS

COMPUTATIONAL SOCIAL SCIENCE, 1

- Exploit insights and theories from social science research
 - Typically small scale, fragmentary studies
 - E.g., word imageability studies in psychology
 - E.g., topic introduction in communication
- Insights converted into algorithms
 - *Simplicity*: break problems into manageable sub-problems
 - *Validation*: make sure people can consistently identify them
 - Low agreement may indicate the problem is ill-defined
 - Training annotators may take many hours
 - *Automation*: Does not need to be 1-1 with human annotation
 - But must show a high degree of correlation $\alpha > 0.7$

COMPUTATIONAL SOCIAL SCIENCE, 2

- Data collection in controlled experiments
 - Maximize signal to noise ratio
 - Ensure that the required phenomena are present
 - Ground truth: survey participants before and after
- Large scale validation
 - Use experimental data first
 - Then move to data in the “wild”
 - Evaluation datasets (e.g., NIST, trained annotators)
 - Ground truth proxies (e.g., Reddit up-votes)
 - Must assure data is representative
- Optimization
 - Use machine learning to optimize parameters

COMPUTATIONAL SOCIAL SCIENCE, 3

- Additional tools needed
 - Annotation tools and manuals
 - Training guides
 - Logging systems
 - Surveys
 - Statistics
 - Experimental protocols, IRB approvals
- Where do we go from here?
 - Validated computational tools replace human coding
 - Apply to large scale datasets
 - But the data must be representative!

WHY NOW?

- Computers are powerful enough to handle “big” complex data
- Human interactions are captured daily on unprecedented scale
 - Social media
 - Online search
 - Online transactions
 - GPS
 - Cell phones
 - Multiplayer online games

WHAT CAN WE LEARN FROM THESE DATA?

- Group level pragmatics
 - Leaders and influencers
 - Power and conflicts
 - Cohesion, factions, alienation
 - Internal dynamics
- Society level knowledge and models
 - Same as group level, plus
 - Information, influence dissemination
 - Opinion trends, prevailing beliefs
 - Movement patterns, demographics
 - Economic, climate, crime, education
 - Conflicts

HOW?

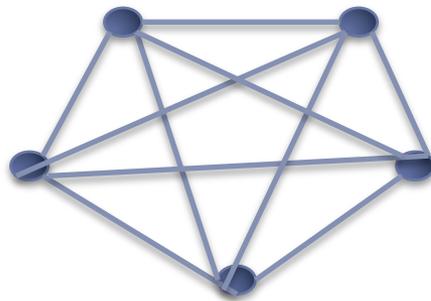
- *Deep* big data analytics
 - Human information is plentiful
 - It is also very complex, noisy, conflicted, incomplete
 - Real patterns often obscured by false ones
- Extracting the meaning
 - Sometimes people tell us what they are doing
 - E.g., Twitter, Facebook
 - More often they do not
 - Intentions, agendas, plans, none of your business
 - Understanding the effects a person's behavior has on others
 - Persuasion, respect, attitudes

WHAT KIND OF DATA?

- Text – readily processed, stored, and plentiful
 - We will focus on this first
- Speech – harder to process, less reliable
 - But: pitch, tone, volume, etc.
- Motion: gestures, movement, facial – indirect communication acts?
 - Requires video signal processing
- Search keywords – ok text, but different
 - I want X, I like X, ...
- Like/Unlike etc. – ok, text but different yet
- Social network: friends, family, colleagues, business
 - Phone, SoMe, shopping, credit cards

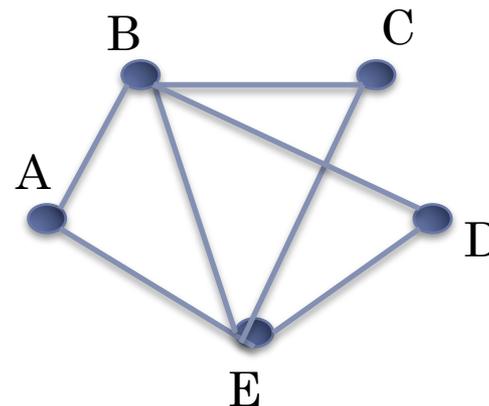
SOCIAL NETWORK

- The geometry of social network matters
 - Who is connected to whom
 - Who hears whom? (note that this is different from who speaks to whom)
 - Is this one-way or two way channel?
- A traditional conversation forms a fully connected bidirectional graph but social connections may look quite different



Everyone knows everyone else

VS.



A, C and D know only B and E
 B, E know everyone

SHAPE OF THE NETWORK

- In networks that are not fully connected
 - Or have one-way connectors
 - Or slower/faster connections
- Communication may be impeded in various ways
 - Less connected members has distorted view of the “world”
 - Highly connected individuals can reach (and influence) more people
 - Highly connected individuals can exploit their connectivity to push unpopular ideas

BIG DATA CAUTIONARY

- Computational Social Science works with massive data
 - Computers can quickly process large quantities
 - A huge advantage over manual process
 - No more worries about sparse data, sample power, significance
- BUT: as with any human data we must consider:
 - Is the data representative?
 - E.g., Twitter users not representative of U.S. population
 - Is there a causation, or merely a correlation
 - If two phenomena co-occur, there may be another cause
 - Remember that with enough data any correlation may achieve significance levels
 - Is the observed phenomenon real and meaningful?
 - Can people agree on identifying it?

SOME SPECIFIC RESEARCH QUESTIONS

- Can machines understand human language?
 - Conversely, can they communicate effectively?
- Can machines understand human behavior?
 - In groups, do social behavior, relations matter?
- How is meaning shaped by behavior?
 - Can machines be taught to behave “socially”?

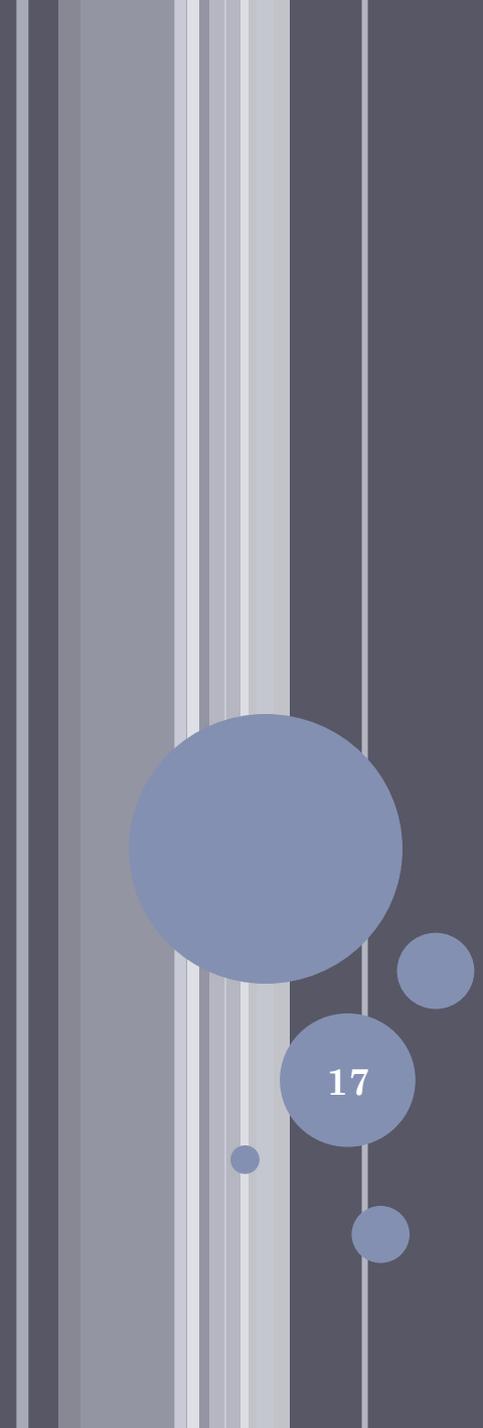
- A lot of this sounds like artificial intelligence (AI)
 - But it is not quite it, or perhaps it’s beyond it.
 - Being intelligent does not mean to be social
 - Anecdotal: Mr. Spock, Cmdr. Data

WHERE DO WE START?

- Language!
 - Plentiful, relatively easy to handle
 - People like to talk, interact
 - Twitter, Facebook, blogs, etc.
- We can extract plenty of information from it
 - Sociolinguistic structure of discourse
 - How this structure shapes opinions and beliefs
 - What is the role of cultural context
 - What are speakers' attitudes, beliefs, and how they change
 - Social roles and relations between participants

ETHICS CONSIDERATIONS

- Before we move on, a big cautionary
- We are handling human generated data
- This means we are dealing with information that may be private and protected under the law
 - Obvious: names, addresses, SSNs, etc.
 - Less obvious: cumulative identification effect
 - Most research requires IRB approval
- Discovery and interventions
 - What if potentially criminal activity is discovered?
 - What if an intervention (e.g., opinion shaping) leads to an unintended effect?

The left side of the slide features a vertical stack of decorative elements. At the top, there are several thin, vertical stripes in shades of gray and white. Below these, a large, solid blue circle is positioned. To its right, a smaller blue circle is placed. Further down, a medium-sized blue circle contains the number '17'. Below this, there are two more small blue circles, one to the left and one to the right of the central vertical axis. The background of the slide is a solid, dark blue color.

SOCIOLINGUISTIC STRUCTURE OF DISCOURSE

17

SOCIAL DIMENSIONS OF DISCOURSE

- Any discourse is a social interaction
 - Task/problem solving
 - Negotiation
 - Debate
 - when ordering food at a restaurant
- More than just the words exchanged; a discourse does:
 - Reflect participants' own goals and opinions
 - Influence other participants' goals and opinions
 - Support or disrupt other participants' objectives
 - Aim at finding out what others think

A DISCOURSE TAKES MANY FORMS

- Face-to-face meetings
- Telephone conversations
- On-line chat, also: on Facebook, Twitter, etc.
- As well as: editorials, scientific publications, and so on



- In almost all cases complex social phenomena are observed...

LANGUAGE REVEALS BEHAVIOR

- Language reflects social phenomena in groups, whenever people interact
 - Joint tasks
 - Co-habitation
 - Games
- Types of social phenomena of interest
 - Social roles
 - Behaviors
 - Intentions
 - Hierarchy
 - Structure
 - Any changes to the above

SOCIAL PHENOMENA IN GROUPS

○ **Leader** is a social role.

- The leader: guides group toward an outcome, manages interaction, controls discussion
- Other group members: recognize the leader



○ **Influencer** is a social role

- The influencer: introduces ideas, has credibility
- Other members: pick up & support influencer's ideas



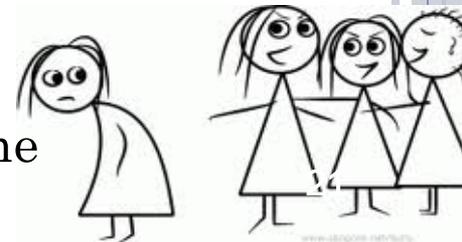
○ **Pursuit of Power** is a social role

- Anyone who makes repeated efforts to gain control
- and whose behavior causes tension within the group



○ **Group Cohesion** is a social state

- Persistence in group objectives or values over time



SOCIAL ROLES AND BEHAVIORS



- How are these accomplished with language?
- How can we detect their presence through language?
 - Words, sounds, and body language in F2F situations

WHAT ARE SOCIOLINGUISTIC BEHAVIORS?

- *Topic Control* – attempts by a discourse participant to impose the topic of conversation
- *Task Control* – an effort by a group member to define the group's goal and/or steer the group towards it.
- *Argument Diversity* – deploying a broader range of arguments in conversation
- *Topical Positioning* – speaker's attitude on main topics (meso-topics) of discussion.
- *Disagreement* – making explicit or implicit utterances of disagreement, disapproval, or rejection.
- *Agreement* – making explicit, unqualified utterances of agreement, approval, or acceptance

AND MANY MORE to be discussed in this course



EXAMPLE: TOPIC CONTROL BEHAVIOR

- **Topic Control:** *attempts by a discourse participant to impact or impose the topic of conversation.*
 - Can be observed over a length of discourse
 - Relative to other participants' behavior.
 - Speakers display varying degrees of Topic Control
- Topic Control correlates with speaker's ***influence*** in discourse
- One way to do it: *introduce topics into conversation that others keep talking about*

TOPIC INTRODUCTIONS → TOPIC CONTROL

First
mention

Turn 42- **LE** (8:33:03 PM): I guess we should just start, not wait for CS and SH?

Turn 43- **JR** (8:33:32 PM): sure

Turn 44- **KN** (8:33:43 PM): ok

Turn 45- **LE** (8:34:02 PM): Fundraising was Mark, Nanny was **Carla**, I think, if you were talking about my comment.

Turn 46- **JR** (8:35:05 PM): gotcha- so that is not he most important to get this job....

Turn 47- **JR** (8:35:23 PM): sorry about my typos- not used to this laptop yet

Turn 48- **JR** (8:36:27 PM): wanna go thru carlas resume first ?

Turn 49- **KN** (8:36:43 PM): sure

Turn 50- **LE** (8:36:44 PM): Sure.

Turn 51- **KN** (8:37:00 PM): i wonder how old carla is

Turn 52- **LE** (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.

Turn 53- **KN** (8:37:30 PM): she's got a perfect driving record and rides **horses!**
coincidence?

Turn 54- **JR** (8:37:35 PM): '06 high school grad

Turn 55- **KN** (8:37:44 PM): i think she rides a **horse** and not a car!

TOPIC INTRODUCTIONS → TOPIC CONTROL

First
mention

Turn 42- **LE** (8:33:03 PM): I guess we should just start, not wait for CS and SH?

Turn 43- **JR** (8:33:32 PM): sure

Turn 44- **KN** (8:33:43 PM): ok

Turn 45- **LE** (8:34:02 PM): Fundraising was Mark, Nanny was **Carla**, I think, if you were talking about my comment.

Turn 46- **JR** (8:35:05 PM): gotcha- so that is not he most important to get this job....

Turn 47- **JR** (8:35:23 PM): sorry about my typos: not used to this laptop yet

Turn 48- **JR** (8:36:27 PM): wanna go thru **carlas** resume first ?

Turn 49- **KN** (8:36:43 PM): sure

Turn 50- **LE** (8:36:44 PM): Sure.

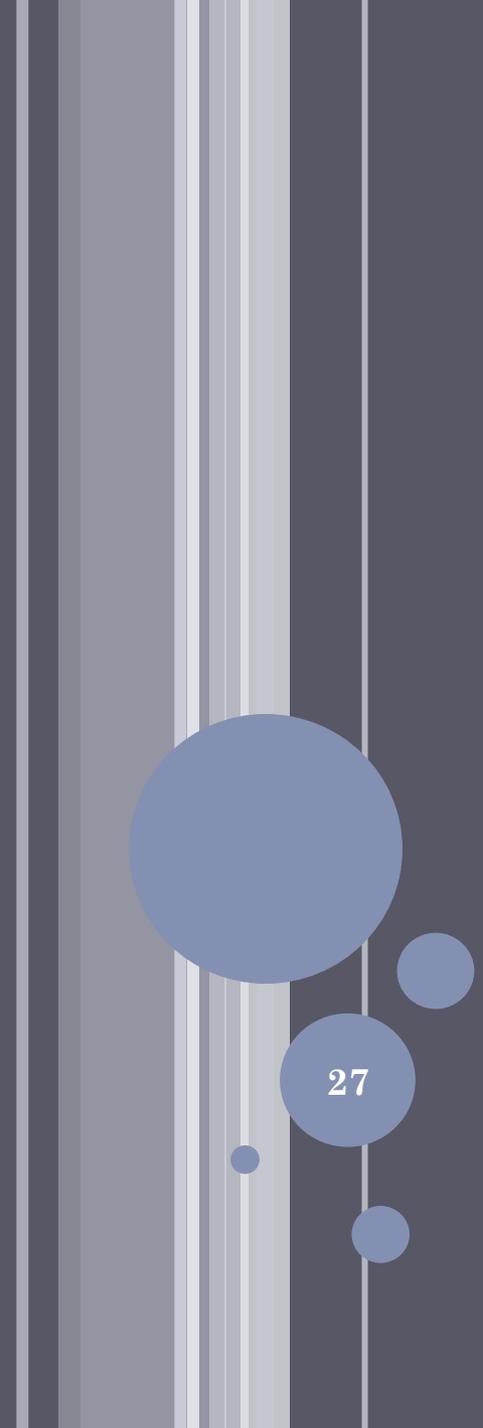
Turn 51- **KN** (8:37:00 PM): i wonder how old **carla** is

Turn 52- **LE** (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.

Turn 53- **KN** (8:37:30 PM): **she**'s got a perfect driving record and rides **horses**! coincidence?

Turn 54- **JR** (8:37:35 PM): '06 high school grad

Turn 55- **KN** (8:37:44 PM): i think **she** rides a **horse** and not a car!

A decorative vertical bar on the left side of the slide, consisting of several thin, parallel lines in shades of gray and blue. To the right of these lines are several overlapping circles of varying sizes, also in shades of blue and gray, creating a modern, abstract design.

BELIEFS AND ATTITUDES

27

PRIVATE STATES

- People hold a variety of opinions and attitudes towards other people, things, and events.
 - Beliefs: opinions that something is true
 - Sentiments: opinions that something is positive or negative (right/wrong, etc.)
 - Other attitudes: being for/against, important/not
 - Emotions: more refined sentiments and attitudes
- Private states are not fixed and can change
 - Can change over time in response to events
 - Influential people can persuade others to change their private states

JOE'S BELIEF SYSTEM ABOUT SEN. SMITH

Belief statement

How strongly held

Joe's attitude towards

| | b_i | e_i | $b_i e_i$ |
|-------------------------|-------|----------------|-----------|
| supports defense cuts | +3 | -2 | -6 |
| helpful to constituents | -3 | +3 | -9 |
| respected in Senate | +2 | +1 | +2 |
| unethical | -2 | -3 | +6 |
| | | $\sum b_i e_i$ | -7 |

CULTURE LEVEL ENCODED ATTITUDES

- Language provides mechanisms to capture attitudes prevalent in a culture
 - Words, phrases are often associated with polarity
 - Affective norms, e.g. *dog* could be positive or negative
 - Figurative language: mapping one concept on another
 - Metaphors encode attitude and affect towards concepts
 - E.g., “*fall into poverty*” compares poverty to a deep hole with negative affect.
- Use of figurative language can convey beliefs, attitudes, and emotions
 - Can be used to form society level analyses

U.S. GUN DEBATE: TOP LEVEL VIEW

- How each side views the Guns Issue

Government oversight
(GOVTO, Anti-gun)

Individual oversight
(INDO, Pro-gun)

DISEASE



WAR

