



Text to Networks and Semantic Networks



Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

 Institute for Software Research


 Carnegie Mellon


Center for Computational Analysis of
Social and Organizational Systems
<http://www.casos.cs.cmu.edu/>



From Texts to Networks From AutoMap/NetMapper to ORA

- Extract a network from texts using a text mining tool
 - E.g. AutoMap or NetMapper
 - Easy approach is D2M – data-to-model
- Load relational data into ORA
 - Visualize the network
 - Generate specialized reports
 - Hot topics report
 - Semantic network report
 - Generate other reports



 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 2



Carnegie Mellon

**Network Text Analysis:
two-step, iterative process**

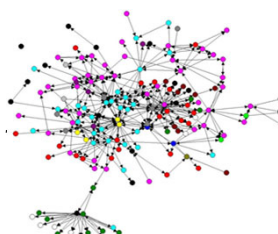
1. Text pre-processing:

- Condense data to terms relevant for domain and research question
 - Normalization
 - Reduction
- Supports the task of finding meaningful interpretations of texts

Natural Language Processing and Information Extraction precondition for NTA

2. Link Formation:

- Determine how the relevant concepts will be linked into ties



CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 3

Carnegie Mellon

IST Institute for Software Research

Data To Model Procedure

- Preprocessors
 - Clean and compartmentalize texts
 - Speeds processing
- D2M script
 - Common bi-grams
 - Thesauri
 - Deletes
 - Relation extraction
 - Ontology cross-classification
- Postprocessing
 - Attribute addition
 - Geo-location
 - Membership and belief inference
 - Event inference (still in development)

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 4



Carnegie Mellon
IST Institute for Software Research

Key: Coding Choices


- Text pre-processing:
 - Condense data to concepts relevant for domain and research question
 - Supports the task of finding meaningful interpretations of texts
- Statement Formation:
 - Determine how the relevant concepts will be linked into statements
 - Analysts: tailor them to research question

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 5

Carnegie Mellon
IST Institute for Software Research

Text pre-processing: Filters



- Deletion
 - remove non-content bearing words such as determiners, conjunctions, articles, prepositions
 - Highly efficient
- Thesauri
 - Convert diverse concepts to the same concept e.g., aliases
 - Enables theme construction


CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 6




Carnegie Mellon
IST Institute for Software Research

Text pre-processing: Apply Background Information



- Thesauri
 - A thesaurus converts from one concept to another
 - Column N: relevant text-level concepts in the
 - Column N+1: corresponding higher-level concepts
 - Represents text-level concepts in a more abstract or general way
 - a single higher level concept typically has multiple entries in the thesaurus.
- Translation
 - Translations convert from other languages to English
- Attributes
 - Known attributes of the concept
 - E.g. for people this might include first name, gender, role
 - E.g. for locations this might include country



IST Institute for Software Research

CASOS


June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 7

Carnegie Mellon
IST Institute for Software Research

Universal and Domain

Terminology in information fields might depend on data and research question that is highly specialized

- Universal
 - Thesauri – General conversions that most people want
 - Delete List – General stop words, numbers, measures and temporal indicators
- Domain
 - Thesauri – Special conversions for the specific data and research question not covered by universal or that alter the universal choice
 - Delete List – Special concepts that are uninteresting in the domain
 - E.g. all concepts with low TFIDF or low frequency or isolates
- Thesauri's over-ride delete lists
- Domain thesauri over-rides universal thesauri
- Universal delete list over-rides domain delete list



CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 8

Carnegie Mellon
IST Institute for Software Research

Support Tools for Creating Domain Files

- Thesaurus creation support:
 - Union Concept List
 - N-grams (especially bigrams)
 - Feature Selection (e.g. tf-idf)
 - External sources (CIA World Fact Book, WordNet)
 - Bootstrapping
- Translation
 - People
 - Online word lists
 - Google translate
 - Other online translators
- Attributes
 - Coding schemes
 - Online structured data

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 9

Carnegie Mellon
IST Institute for Software Research

Pre-processing

- “John Doe is actively involved in several industry and civic associations.”
- Deletion:
 - Remove non-content bearing concepts
 - High data reduction rates
 - Delete List: is, in, and
 - Delete List creation support: concept list per text and corpus (Zipf’s law)
 - Adjacency:
 - Direct: John Doe actively involved several industry civic associations.
 - Rhetorical: John Doe xxx actively involved xxx several industry xxx civic associations.
- Zipf, George Kingsley. 1949. Human Behavior and the Principle of Least Effort. Cambridge, Mass.: Addison-Wesley.

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 10




Carnegie Mellon
ISR Institute for Software Research

Pre-processing

- Stemming
 - Detect inflections and derivations of concepts
 - Convert each concept into related morpheme
 - John Doe is actively involved in several industry and civic associations.
 - Porter: John Doe be activ involv in sever industri and civic associ.
 - Krovetz: John Doe is actively involve in several industry an civic association.

Porter, M.F. 1980. An algorithm for suffix stripping. I 14 (3): 130-137.
Krovetz, Robert (1995). Word Sense Disambiguation for Large Text Databases. Unpublished PhD Theis, University of Massachusetts.


 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 11

Carnegie Mellon
ISR Institute for Software Research

Pre-processing

- Thesauri
 - Convert concepts with more abstract concepts (generalization thesaurus) or meta-network entities (meta-network thesaurus)
 - John Doe is actively involved in several industry and civic associations.

• Generalization thesaurus	meta-network thesaurus
• John Doe/ John_Doe	John_Doe/ agent
• involved/ engage	engage/ task
• Industry/ economy	economy/ resource
• civic associations/ community	community/ organization

 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 12



Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Getting Started - Data Collection

- Newspaper articles
- Books
- Journal articles
- Abstracts
- Interviews
- Essay Questions
- Email
- Web pages
- Dictionary entries
- Scripts
- Concept circles
- Power point presentations

Any type of raw text

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 13

Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Representation Scheme

DNA Term	Term	Graphic	Equivalent Ideas
Node Link	Concept Link		Word Phrase
Dyad	Statement		Sentence Implication Relation Idea Fact
Network	Map		Definition Mental Model Frame

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 14



Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Concept Circle

Idea

- List of concepts
- Place concepts in circle
- Have respondent draw lines between linked concepts

When Appropriate

- Want to cue respondent
- Want only predefined concepts
- Less than 50 concepts

Usage Hint

- Train respondents on small task (10 concepts)
- Go over map with respondent

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 15

Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Concept Circle - Example

Clustering Task # 2: April 26, 1989 Name _____

Directions: These words have been mentioned in class lectures over the past semester. Please draw a line between pairs of words which you believe should be connected. It is important that all connections that you intend to make be clear and easy to see. Please do not draw so many lines on any one worksheet that you cannot easily see how you've connected those words.

analysis approaches
 agent argument
 action aspects
 abstract attribution
 abstract author(s)
 XbutY authority
 writing background
 weaknesses tree body
 synthesis choose the solution
 support citation
 summary community
 strengths contribute
 sources converge
 solution define the problem
 situation design
 shared differences
 similarities directions
 seeing the issue disclaimer
 return path elaboration
 result explore
 relevant faulty path
 reader focus
 qualification framework
 progress goal
 problem definition historical account
 problem case(s) incompatible
 problem introduction
 positions irrelevant
 plan issue
 perspective knowledge
 paradigm case(s) lens on an issue
 paper line of argument
 original main path
 novelty new milestone(s)

Palmquist, Kaufer,
 Carley Learning to
 Write Study 1989

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 16



Carnegie Mellon
IST Institute for Software Research

Concept Circle - Cont.

↗

Variations:

When Respondent Draws Lines

- Place strength on lines
- Place arrows on lines for causality
- Place marker on lines for type of link

Application Process

- Can be applied by interviewer during interview
- Can be done as reading text

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 17

Carnegie Mellon
IST Institute for Software Research

Link formation

- In order to extract semantic network from raw or pre-processed data we need to link the nodes
- Multiple approaches exists:
 - Distance based approach: Windowing:
 - Text Unit
 - Window Size
 - Adjacency

Danowski, J. (1982). A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. In R. Bostrom (Ed.), *Communication Yearbook*, 6: 904-925. New Brunswick, NJ: Transaction Books.

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 18



Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Syntactic Based Link Formation

- Distance based approach, Windowing:
 - Text Unit (text, paragraph, sentence)
 - Window Size (2 to N)
 - Adjacency (direct or rhetorical)
- Thesaurus:
 - leader/ leader; civic_association
 - involved/ engage; industry/ industry;
 - civic associations/ civic_association
- Coding Choices:
 - Sentence, thesaurus content only, rhetorical adjacency, window size sentence
- TEXT: Leader xxx actively involved xxx several industry xxx civic associations.

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 19

Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

From Texts to Networks

- Distance based approach, Windowing:
 - Text Unit (text, paragraph, sentence)
 - Window Size (2 to N)
 - Adjacency (direct or rhetorical)

Leader	xxx	actively	involved	xxx	several	industry	xxx	civic	associations.
--------	-----	----------	----------	-----	---------	----------	-----	-------	---------------

Relational data:
no thesaurus.
WS 3:
Leader - actively
actively - involved
Involved - several

Thesaurus content only, direct adj. WS 3 or rhet. adj. WS 10:

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 20



Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH


Extracted Network as Mental Models

- Semantic Network (Map)
 - Representation of a sample of what people have in their minds and use to make sense of their surroundings.
 - Cognitive construct that reflect the subjects' knowledge and information about a certain topic.

Johnson-Laird, P. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20, 403-437.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box; prospects and limits in the search for mental models. *Psychological Bulletin*, 100, 349-363.

 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 21

Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH


One-mode networks: semantic networks

1. One mode networks: semantic networks, cognitive/ mental models:
 - Represent reality that people have in their minds and use to make sense of their surroundings.
 - Cognitive constructs that reflect the subjects' knowledge and information about a certain topic.
- Applications:
 - Interviews: individual and group-level
 - Self-presentations: annual reports, self-applications, mission statements
 - Groups: (Shared) Situational Awareness

Johnson-Laird, P. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20, 403-437.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box; prospects and limits in the search for mental models. *Psychological Bulletin*, 100, 349-363.

 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 22



Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

One-mode networks: Example

Diesner, J., Kumaraguru, P., & Carley, K.M. (2005). Mental Models of Data Privacy and Security Extracted from Interviews with Indians. 55th Annual Conference of the International Communication Association (ICA). New York, NY, May 26-30, 2005

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 23

Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

Hints

- If using exact words
 - use a frequency level – e.g., only if used more than "3" times, or used by 10% people
- If generalizing
 - create thesaurus file with examples of how to code words, be very general
- General
 - code 2 or 3 texts first to make sure you like set up
- Either code for specific concepts or use delete lists
 - Delete non content bearing words
 - Delete off topic words
- Comment
 - there will be a limited number of words per topic/community – usually 2-3000
 - Some limit to most frequent – this can be misleading
 - May contain off topic
 - Rank may be very different if you use thesauri as key concepts generally have more alternative forms than do other words

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 24



Carnegie Mellon
ISR Institute for Software Research

Illustrative Applications


- Enron
 - Diesner Carley
- Al Qaeda
 - Carley
- Decision Making
 - Carley
- Learning to Write
 - Palmquist PhD
- Risk Perception
 - Bostrom PhD
- World Bank & Poverty
 - Saburi PhD
- Authorship
 - Lewis
- Impact of Email on Invisible College
 - Gardner & Carley
- Robotics in Science Fiction
 - Student Project
- What Children Remember
 - Cicourel
- COVID-19
 - This institute

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 25

Carnegie Mellon
ISR Institute for Software Research

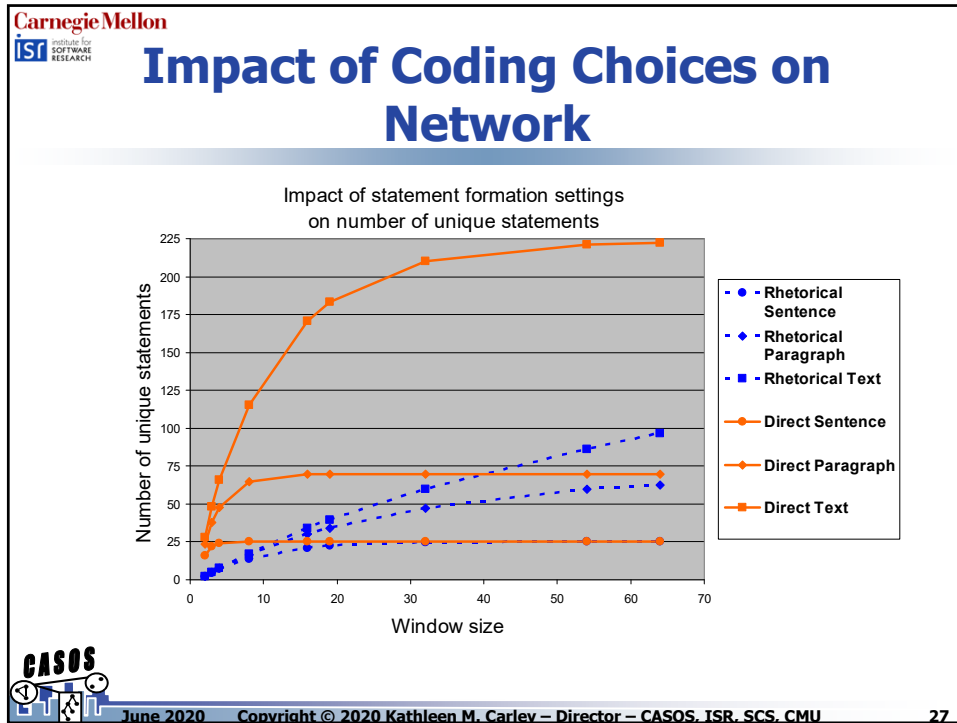
Text pre-processing: Negative Filters



- Deletion
 - remove non-content bearing words such as determiners, conjunctions, articles, prepositions
 - Highly efficient
- Adjacency:
 - Direct: drop delete words and move word right and left to delete word directly next to each other
 - Rhetorical: insert a placeholder where a delete word was removed in order to maintain original distance of all words

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 26



Carnegie Mellon
IST Institute for Software Research

Semantic Network as Meta-Networks

- We might also want to know about other nodes
 - Who is linked to what organizations
 - Who has access to what resources
 - Who possesses what knowledge
 - Who is involved in what tasks
 - Who was at what event
 - What resources are needed for what tasks
 - ...
- We might also want to know about other attributes
 - Who has what personal characteristics
 - What types of locations exist
 - ...

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 28




Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

Semantic Network as Meta-Networks

- Implement meta-network model as ontology : Meta-Network Text Analysis
 - Extract structure of social and organizational systems from corpora
 - Provide relational data that can be further analyzed with other tools and methods
- Use your own ontology
 - Rename existing categories

Diesner, J., & Carley, K.M. (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis. In V.K. Narayanan & D.J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations*, (pp. 81-108). Harrisburg, PA: Idea Group Publishing.


 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 29

Carnegie Mellon
 ISI Institute for SOFTWARE RESEARCH

Multi-mode Networks (aka Meta-Networks)

- Social Networks
 - Supports asking traditional/ baseline questions:
 - Who is connected to whom?
 - What type of relationship?
- With meta-network data we can move beyond traditional questions and also ask:
 - Which agent or group is located where, has access to what resources, possesses what knowledge, is involved in what tasks, has what personal characteristics, ...?
- AutoMap:
 - Use meta-network as is
 - Use self-modified meta-network
 - Use your own ontology
- NetMapper
 - Use meta-network as is
 - Use category 1 to specify your ontology then switch in ORA
 - Relabel categories in ORA

Diesner, J., & Carley, K. M. (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research* (pp. 81-108). Harrisburg, PA: Idea Group Publishing.

 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 30



Carnegie Mellon

Extraction of Relational Data From Texts – Toy Example

Example from UN News Service (New York), 12-28-2004:
 "Jan Pronk, the Special Representative of Secretary-General Kofi Annan to Sudan, today called for the immediate return of the vehicles to World Food Programme (WFP) and NGOs."

extract relational data

One-mode networks

semantic networks

● Knowledge

multi-mode networks

meta-networks

● Person ● Organization
 ● Location ● Resource

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 31

Carnegie Mellon

Extraction of Relational Data From Texts – Two Approaches


- Any type of network:
 - **Locate/ identify** "relevant" nodes, e.g. knowledge
 - Sparse
- Ontologically coded, socio-technical networks:
 - **Classify** relevant nodes according to some ontology or taxonomy

McCallum, A. (2005). Information extraction: distilling structured data from unstructured text. *ACM Queue*, 3(9), 48-57.
 Bikel, D., M., Schwartz, R., & Weischedel, R., M. (1999). An Algorithm that Learns What's in a Name, *Machine Learning* (Vol. 34, pp. 211-231): Kluwer Academic Publishers.
 Diesner, J., & Carley, K.M. (2009). Conditional Random Fields for Entity Extraction and Ontological Text Coding. Accepted for *Journal of Computational and Mathematical Organization Theory* (CMOT).

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 32



Carnegie Mellon
 Institute for SOFTWARE RESEARCH




Default Ontology

- Named Entities
 - Agent
 - Individual actors
 - Specific – unique often with first and last name - Jamie O'Connor
 - Generic – non-unique and often a role - haberdasher
 - Organization
 - Groups, corporations, populations
 - Specific – unique - IBM
 - Generic – a type - Non-government organization
 - Location
 - Places things can be at
 - Specific – unique with lat and lon or place on map – United States of America
 - Generic – may be at multiple locations – hill
 - Event specific, generic
 - Major happenings that impact groups
 - Specific – occur once – World War I
 - Generic – multiple occurrences - Tornado

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 33

Carnegie Mellon
 Institute for SOFTWARE RESEARCH



Default Ontology

- General Entities
 - Knowledge
 - Branches of knowledge
 - Topics of interest
 - Resources
 - Things that are not purely mental – disease, food, wire
 - Tasks
 - Activities – eat
 - Beliefs
 - “isms” - Catholicism
 - Sentiment – positive, negative
 - Belief statements – right to bear arms

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 34




Carnegie Mellon
 Institute for Software Research

Ontological Text Coding

- Requires mapping of words to categories (*agent, event, knowledge, location, organization, resource, task, attribute*)

1. Explicit mapping (positive filters)
 - Deterministic, incomplete, outdated, spelling variations, typos
 - Exploit available data (CIA world fact book, lists, ...)
2. Regular expressions
 - Identifying relevant strings of text (particular characters, words, or patterns) (JGSoft: RegexBuddy)
 - KWIC (keyword in context) ([A-Za-z0-9] = Alphanumeric chars.)
 - URL's, dates, phone numbers


 McCallum, A. (2005). Information extraction: distilling structured data from unstructured text. *ACM Queue*, 3(9), 48-57.

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 35

Carnegie Mellon
 Institute for Software Research

Thesauri Format

conceptFrom	conceptTo	Ontology	nodeType	Category	Country	First_Na me	Last_Nam e	Gende r	Hon orifi c	Veri fied	Lan gua ge	Acr ony m
mine-clearance absolutism	mine_clearance absolutism	task belief										
parents	parent	agent	generic							1		
Pat McGinn	Pat_McGinn	agent	specific	United_States_ Celebrity_Po of_Ame litician	rica	Pat	McGinn	Male		1		

 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 36



Carnegie Mellon
IST Institute for Software Research

Window Based Location of Links

- John_Doe is actively engage in several economy and community.

Map

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 37

Carnegie Mellon
IST Institute for Software Research

Window Based Location of Links with Ontology

- agent is actively task in several resource and organization.

Map

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 38



Carnegie Mellon
 Institute for Software Research

Example Text Files

Hisham Al Hussein
 ... the Philippine government booted the second secretary at Iraq's Manila embassy, Hisham Al Hussein, on February 13, 2003, after discovering that the same mobile phone that reached his number on October 3, 2002, six days later rang another cell phone strapped to a bomb at the San Roque Elementary School in Zamboanga.

Abu Madja and Hamsiraji Ali
 That mobile phone also registered calls to Abu Madja and Hamsiraji Ali, leaders of Abu Sayyaf, Al Qaeda's Philippine branch.

Abdurajak Janjalani
 It was launched in the late 1980s by the late Abdurajak Janjalani, with the help of Jamal Mohammad Khalifa, Osama bin Laden's brother-in-law.

Hamsiraji Ali
 ... Hamsiraji Ali, an Abu Sayyaf commander on the southern island of Basilan, bragged that his group received almost \$20,000 annually from Iraqis close to Saddam Hussein.

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 39

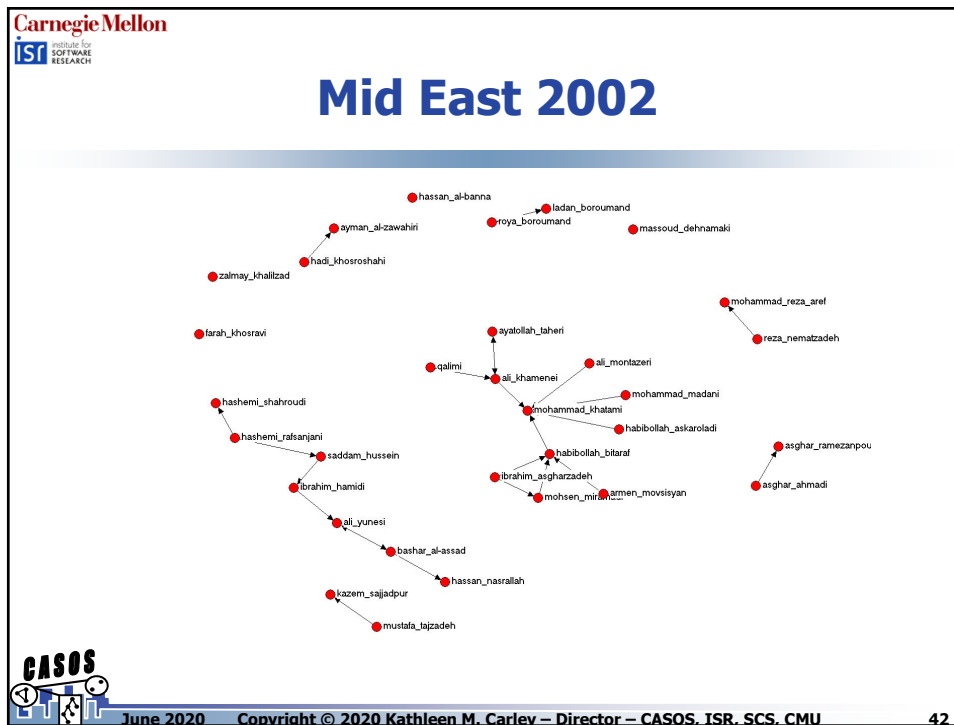
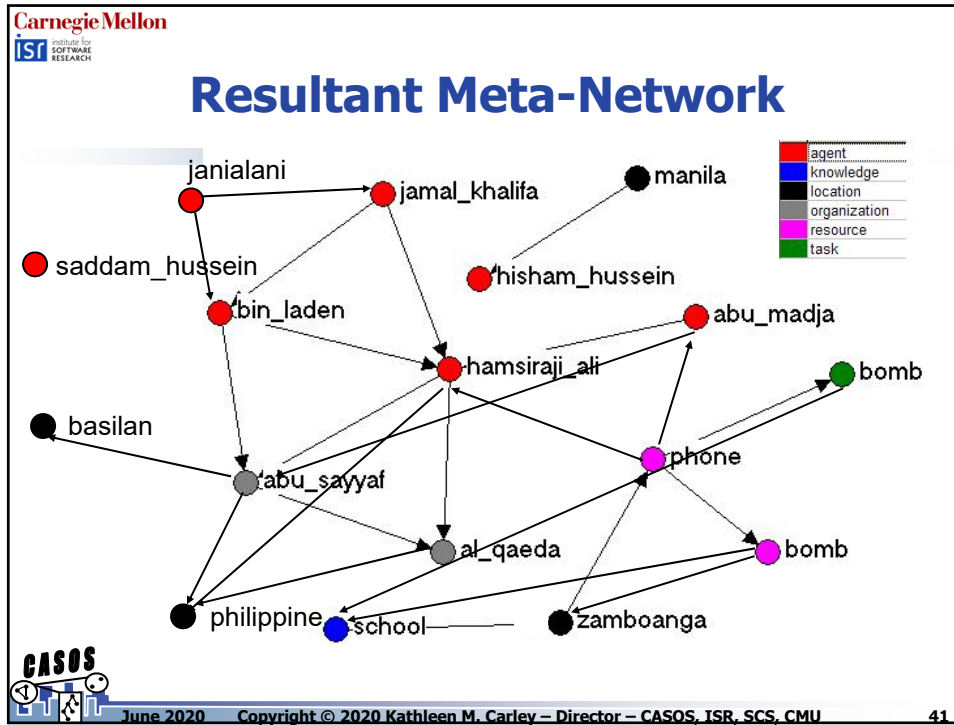
Carnegie Mellon
 Institute for Software Research

Place Concepts in Meta-Network

Agents	Knowledge	Resources	Tasks	Locations	Organizations
abu_madja	school	bomb	bomb	basilian	al_qaeda
bin_laden		phone		philippine	bu_sayyaf
hamsiraji_ali				manila	
hisham_hussein				zaboanga	
janialani					
jaml_khalifa					
saddam_hussein					

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 40





Carnegie Mellon
IST Institute for Software Research

Semantic Network

- Network of concepts
- Sometimes referred to as
 - a mental model
 - An information network
- Generally extracted from texts
- Variations on nodes
 - Words –vs- concepts
 - topics –vs- concepts
- Features of links
 - Co-use or logic or semantic relations
 - Strength – generally represents frequency
 - In text or across texts

CASOS
June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 45

Carnegie Mellon
IST Institute for Software Research

Illustrative Semantic Network

```

graph TD
    Healthcare --> User needs
    Healthcare --> Accounting Systems
    User needs --> Accounting Systems
    User community --> Accounting Systems
    Sale techniques --> Accounting Systems
    Accounting Systems --> Gyms
    Accounting Systems --> Product information
    Accounting Systems --> New products
    Gyms --> Computer capabilities
    Computer capabilities --> Network features
    New products --> Programming
    Programming --> Product problems
    Product problems --> Inventory Systems
    Product information --> Inventory Systems
  
```

CASOS
June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 46



Carnegie Mellon
 Institute for Software Research

Why compare texts as networks: Locating Common Ground

- Do people use same words
- Do people use same words in same way
- Do people evoke the same concepts
- Do people link the same concepts in the same way
- Are the same disciplines employed
- Are the same journals, trade magazines, etc. read

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 47

Carnegie Mellon
 Institute for Software Research

Information Network: Text & Graphics

Student A: I found that scientists engage in research in order to make discoveries and generate new ideas. Such research by scientists is hard work and often involves collaboration with other scientists which leads to discoveries which make the scientists famous. Such collaboration may be informal, such as when they share new ideas over lunch, or formal, such as when they are co-authors of a paper.

STUDENT A

Student B: It was hard work to research famous scientists engaged in collaboration and I made many informal discoveries. My research showed that scientists engaged in collaboration with other scientists are co-authors of at least one paper containing their new ideas. Some scientists make formal discoveries and have new ideas.

STUDENT B

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 48



Carnegie Mellon
IST Institute for Software Research

Network Approach to Communicative Power

- Three key communicative dimensions
 - Intensity/consensus – weighted degree or number of graphs of individual's mental models or number of reports the tie is present in
 - Conductivity – indegree*outdegree or betweenness
 - Density – degree
- Most words will be low on all three dimensions
- Words that stand out on any one dimension will have extra communicative power
- Messages containing such words will be more persuasive

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 49

Carnegie Mellon
IST Institute for Software Research

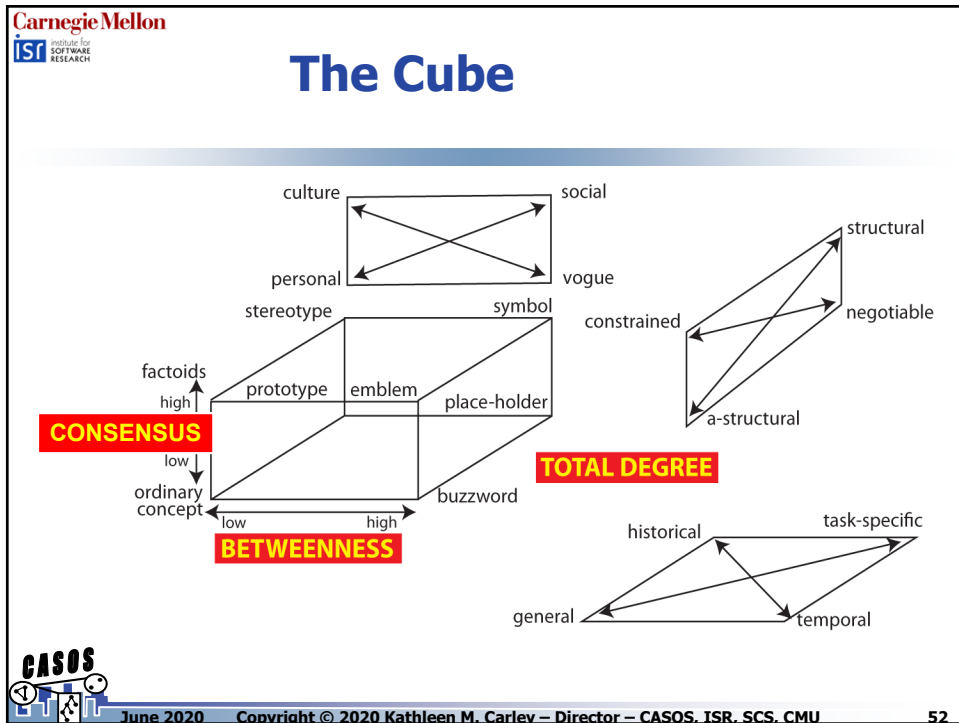
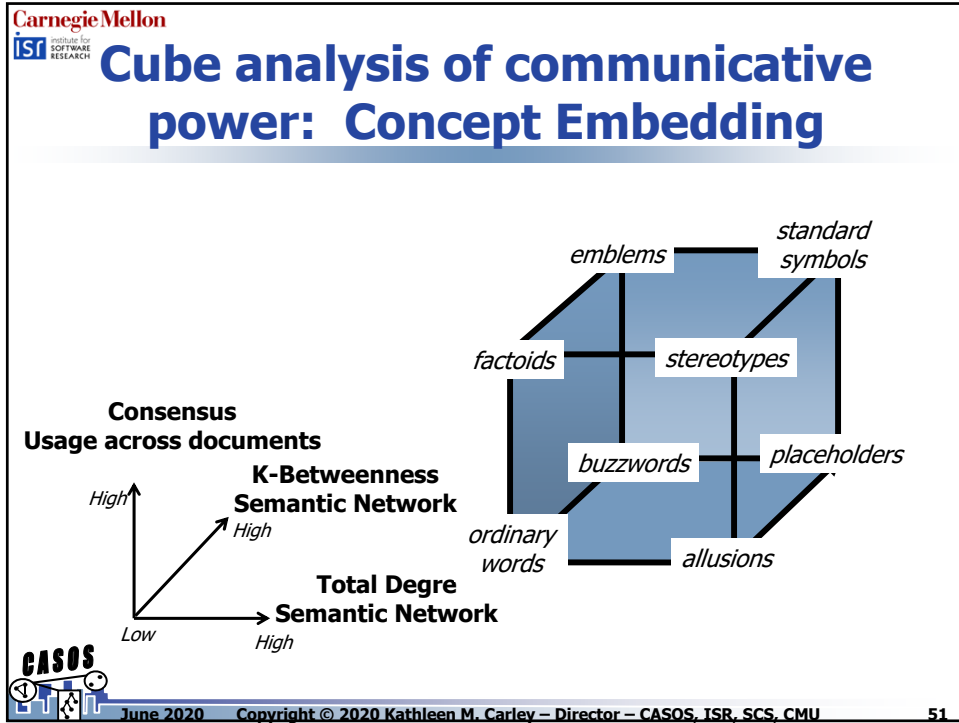
Distributional Information

- Communicative power is a function of reach
- Reach depends on the network position of author and of concept
- Every concept has a position in the meta-network
 - Document by concept – link is number of times document uses concept
 - Concept by concept – link is number of documents that use both concepts (i.e., the author's/document's semantic network)
- Concepts vary in their meta-network position
- Network position of concept determines its communicative power

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU





Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

General: Ordinary Word

sense of frustration

expansion

contradictory evidence

maddeningly

CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 53

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Factoid

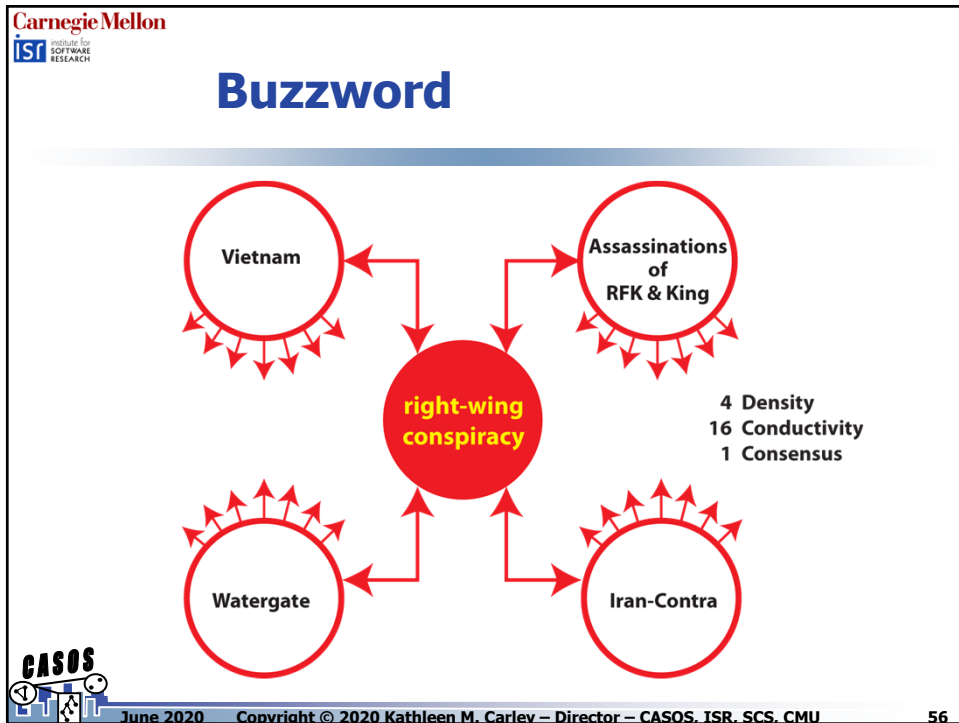
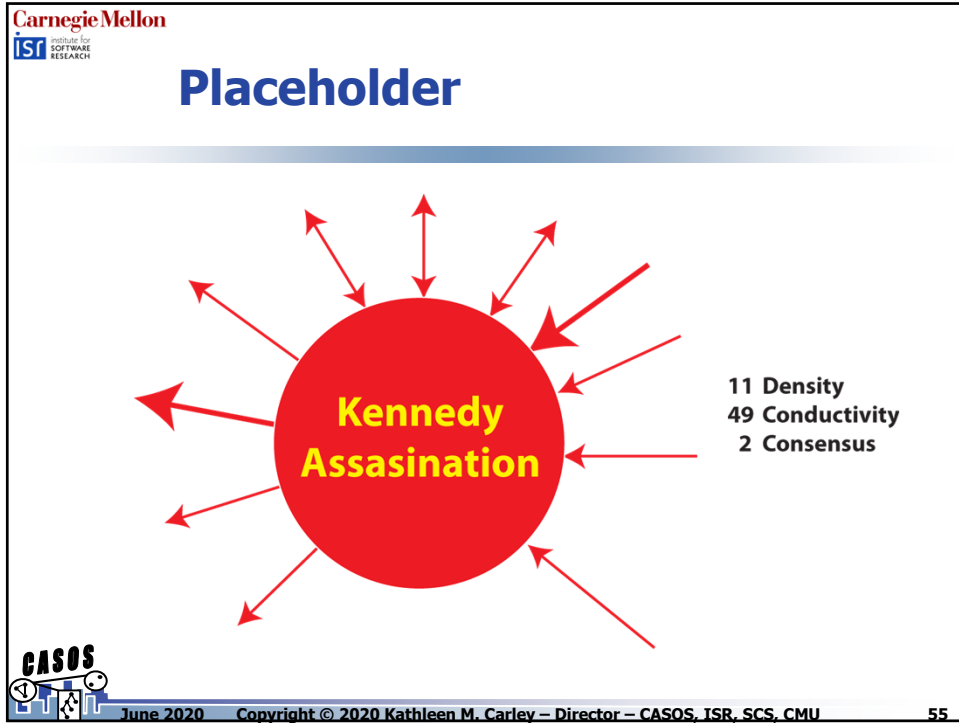
4 Density
3 Conductivity
4 Consensus

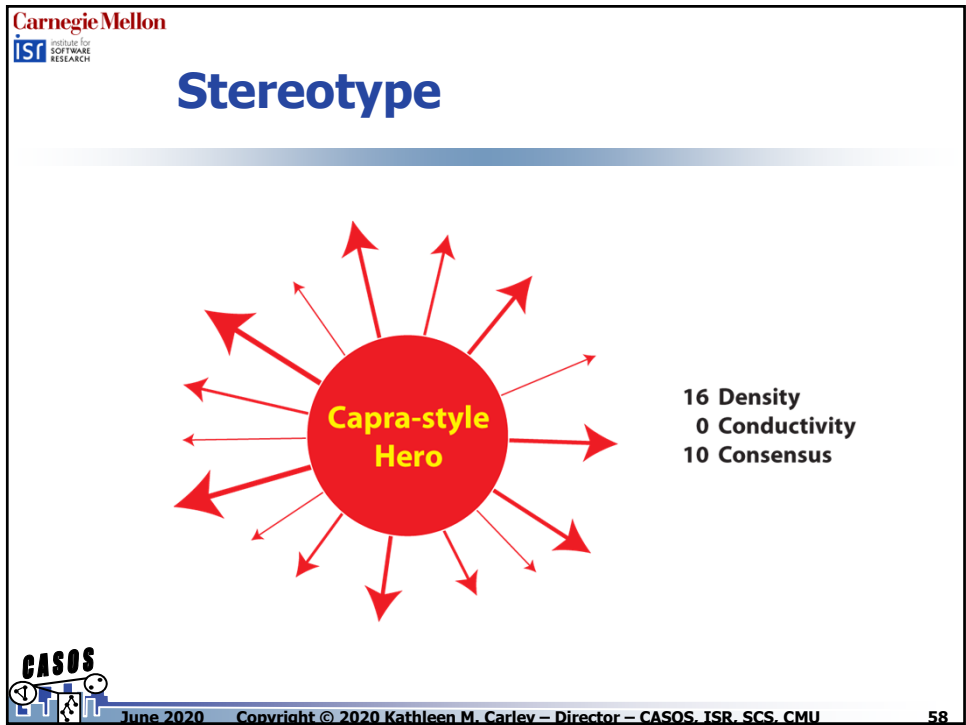
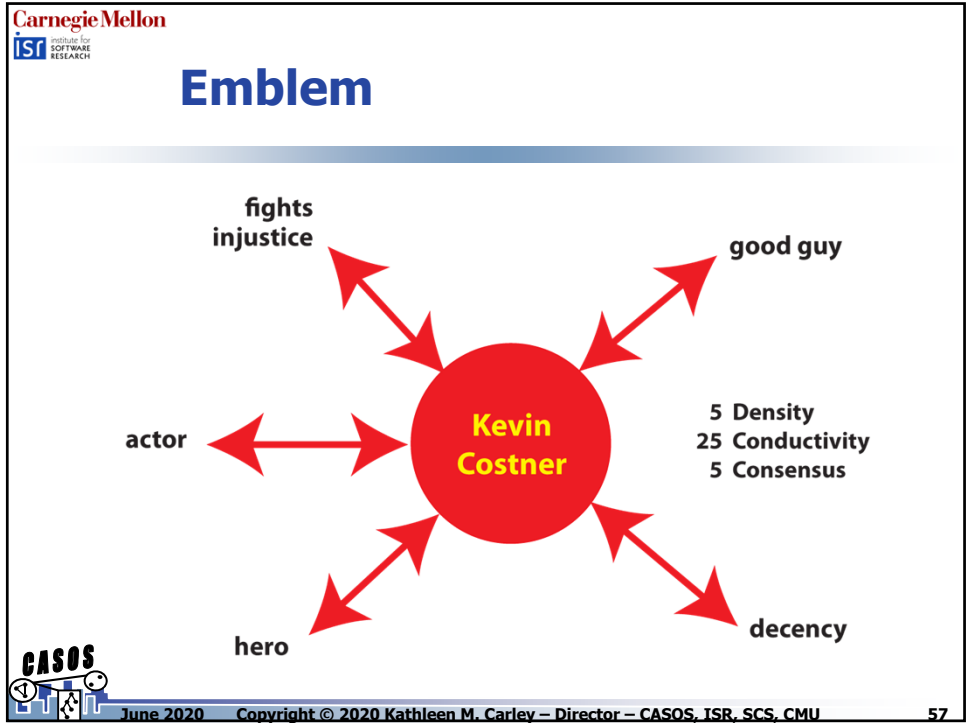
Nov. 22, 1963

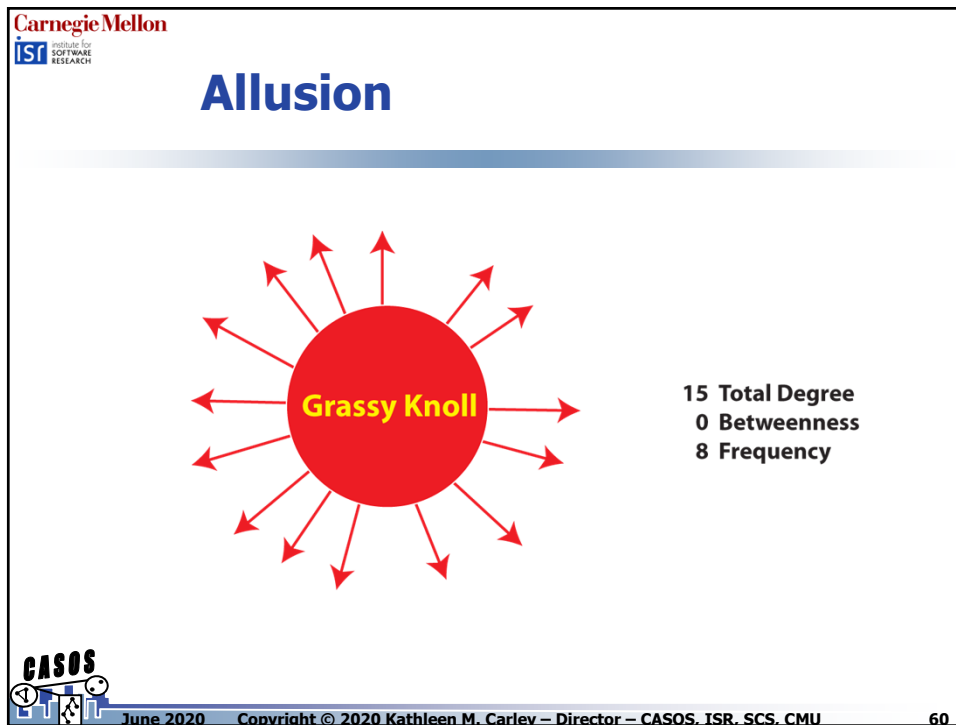
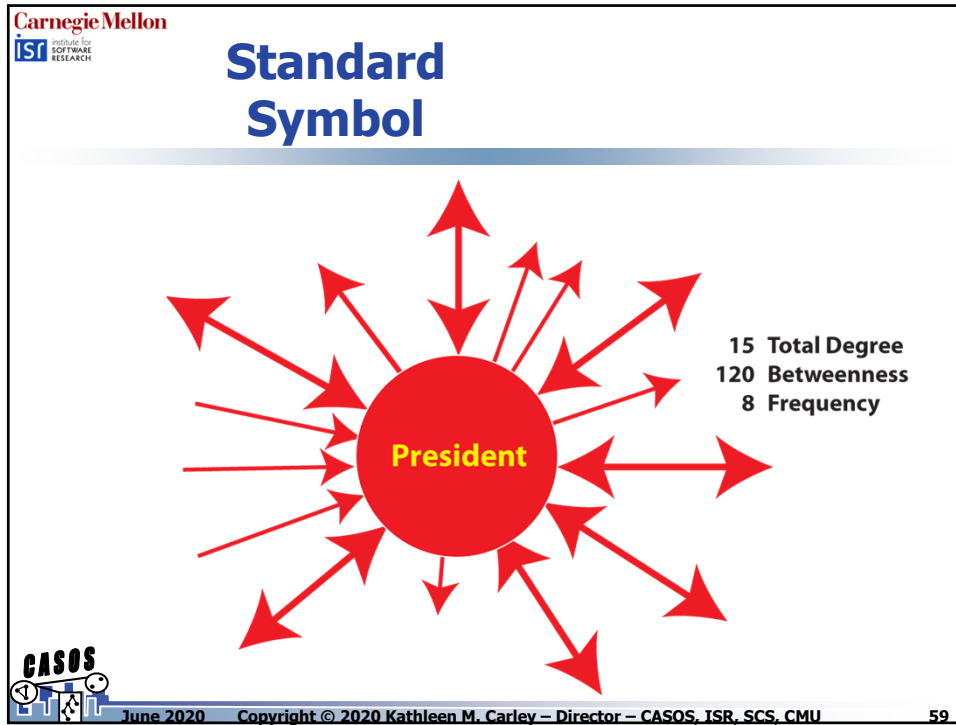
CASOS

June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 54









Carnegie Mellon
 Institute for SOFTWARE RESEARCH

Cube

Purpose	Output
Generate network measures on concept position	Concept by value matrix
Have Ready	
<ul style="list-style-type: none"> • A semantic network • A meta-network that is the union across documents <ul style="list-style-type: none"> - Document x concept - Concept x concept - Often we treat document as location - Concept are type knowledge 	Statistical Package

CASOS June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 61

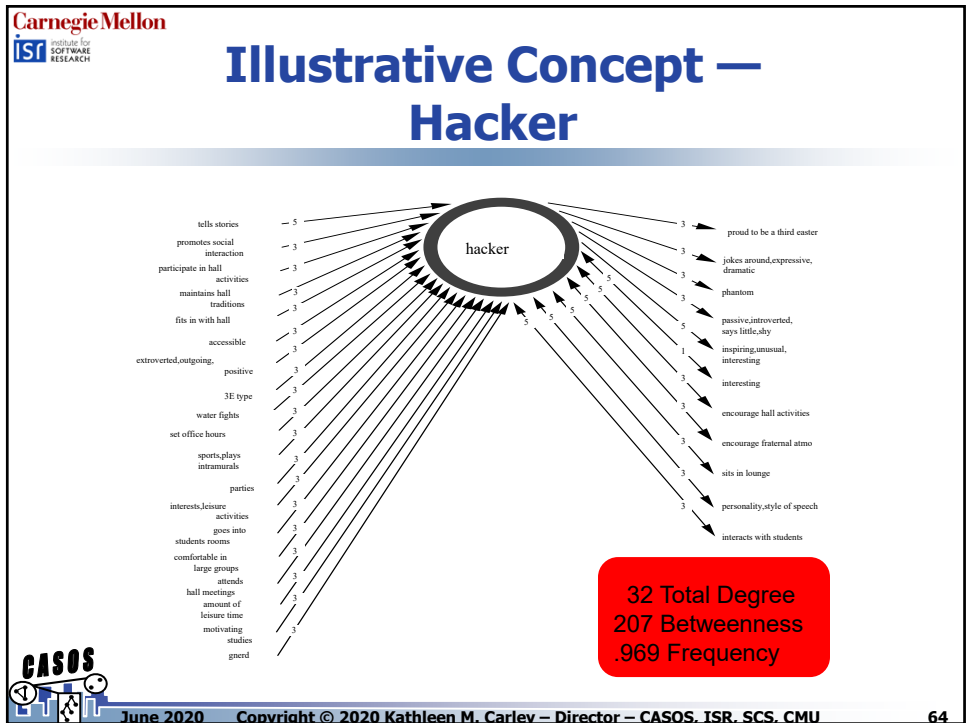
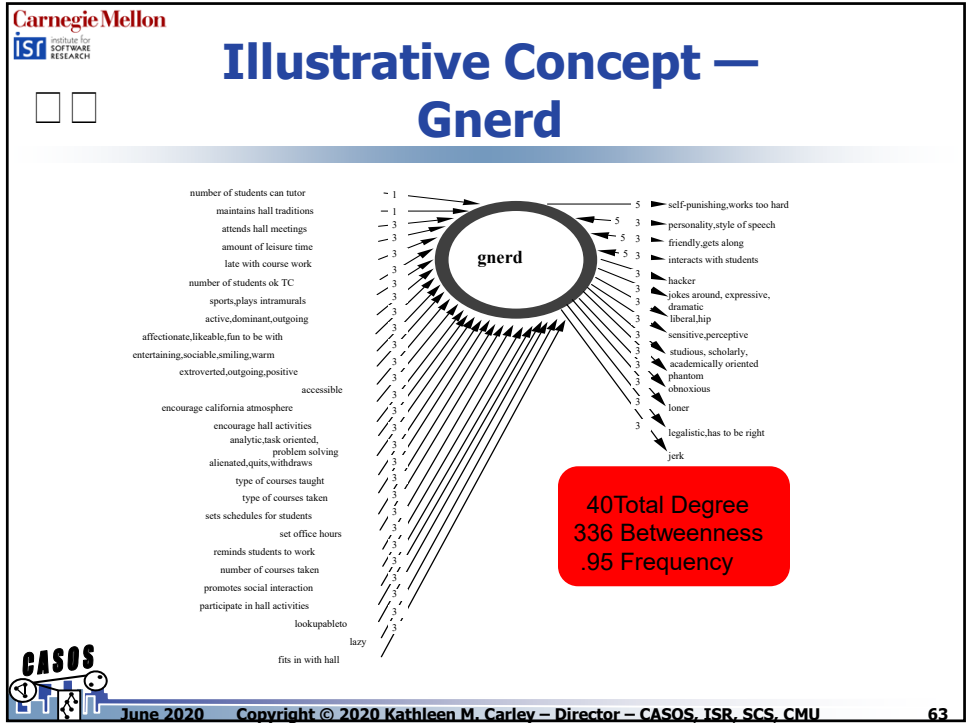
Carnegie Mellon
 Institute for SOFTWARE RESEARCH

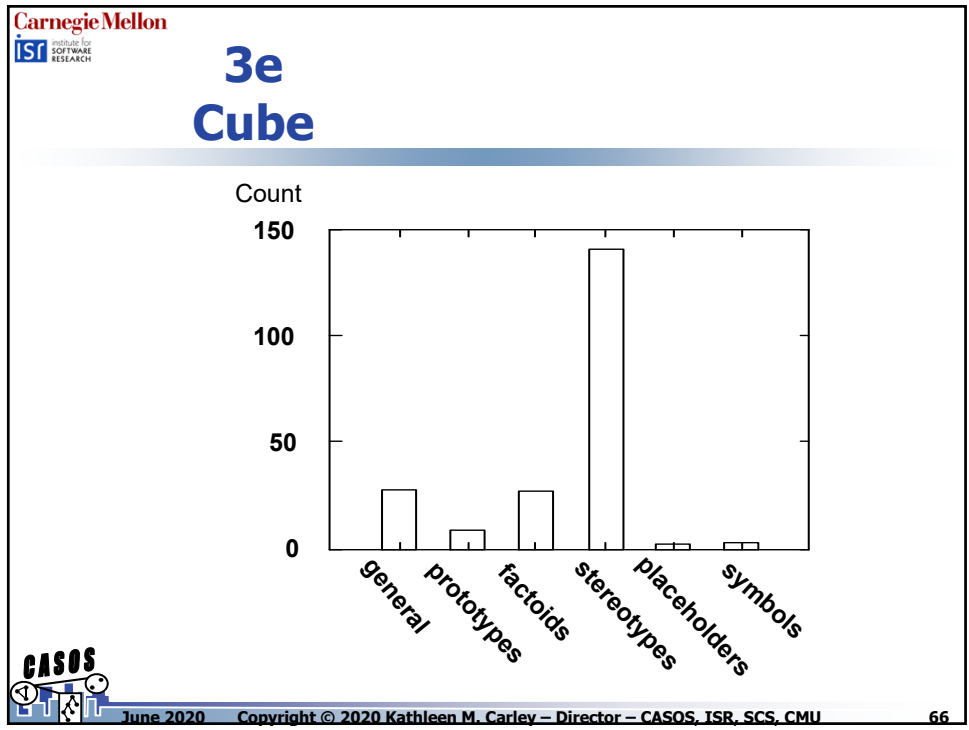
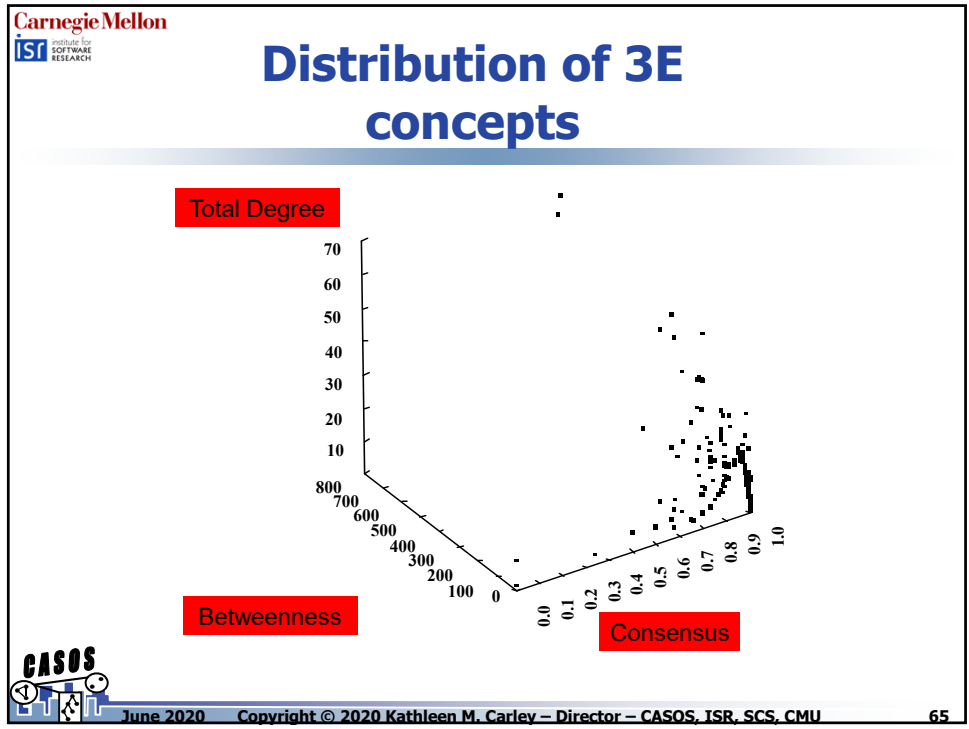
Illustrative Concept - Car

5 Total Degree
6 Betweenness
1 Frequency

CASOS June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 62







Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Highly Communicative Concepts

Degree Betweenness Consensus

" car"	5.000	6.000	1.000	factoids
" kids"	5.000	6.000	1.000	factoids
"intelligence"	26.000	120.000	0.846	prototypes
"interesting"	28.000	132.000	0.643	prototypes
"phantom"	48.000	176.000	0.938	stereotypes
"hacker"	32.000	207.000	0.969	stereotypes
"participates"	32.000	220.000	0.969	stereotypes
"teaching-abl"	36.000	224.000	0.889	stereotypes
"hall-member"	32.000	231.000	0.969	stereotypes
"expert"	31.000	238.000	0.968	stereotypes
"lookupableto"	51.000	270.000	0.882	symbols
"gnerd"	40.000	336.000	0.950	symbols
"accessible"	41.000	378.000	0.925	symbols
"gets-along"	64.000	735.000	0.781	placeholders
"fits-in"	67.000	780.000	0.821	placeholders

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 67

Carnegie Mellon
ISR Institute for SOFTWARE RESEARCH

Cube and Social Knowledge

insists on quiet
 (factoid)
 (10,24,1.0)

mellow
 (factoid)
 (19,70,0.947)

fits-in
 (placeholder)
 (67,780,0.821)

tolerant
 (ordinary)
 (23,76,0.870)

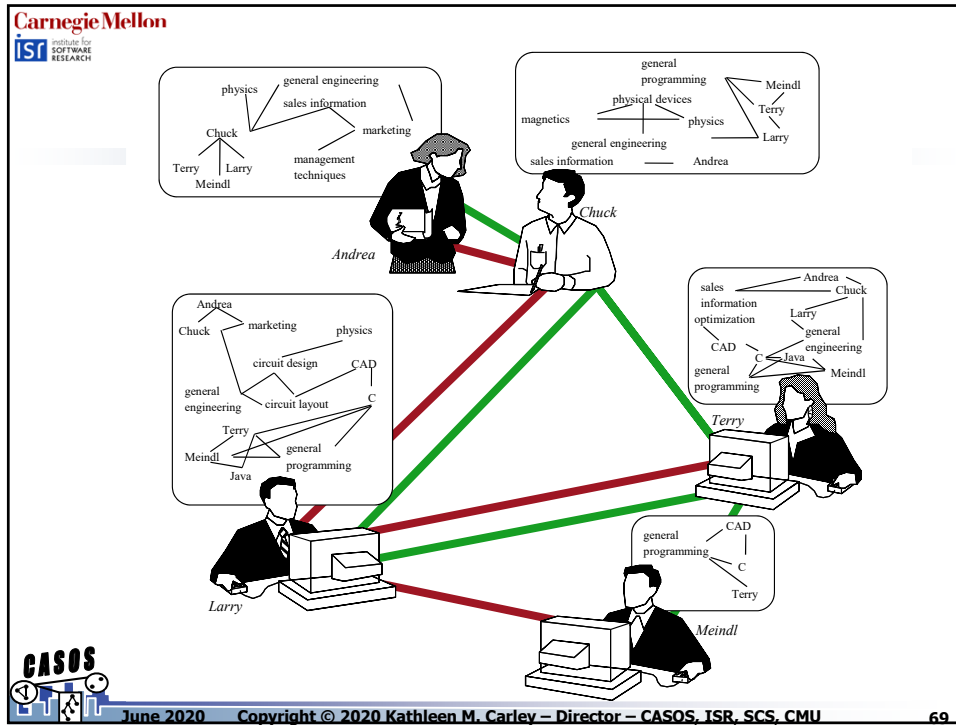
gets-along
 (placeholder)
 (64,735,0.781)

number students ok TC
 (factoid)
 (14,33,0.929)

egocentric
 (ordinary)
 (10,16,0.800)

CASOS
 June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 68





Carnegie Mellon Institute for Software Research

Summary of Cube analysis

- Texts can be coded as networks
- Networks represent the mental model
- Coding texts as mental models focuses on meaning
- Concepts have position in texts
- Examining positions focuses on communicative power
- Classifying concepts creates Meta-Networks

CASOS
June 2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU

