Exploration of News Corpora

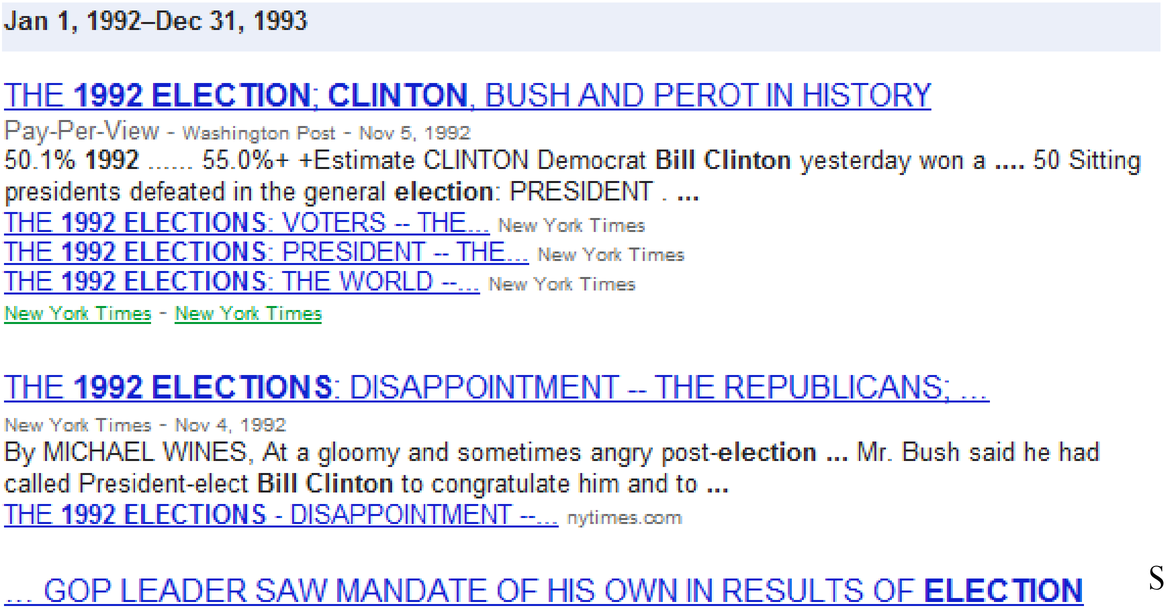
By studying actor evolution &
graph visualization over time

by
Ravee Malla
2008CS50224

under the guidance of
Dr. Maya Ramanath
Dr. Amitabha Bagchi

Motivation

Information Need: ***How did Bill Clinton win the 1992 elections?***



Jan 1, 1992–Dec 31, 1993

THE **1992 ELECTION**; **CLINTON**, BUSH AND PEROT IN HISTORY

Pay-Per-View - Washington Post - Nov 5, 1992

50.1% **1992** ...... 55.0%+ +Estimate CLINTON Democrat **Bill Clinton** yesterday won a **....** 50 Sitting presidents defeated in the general **election**: PRESIDENT . **...**

THE **1992 ELECTIONS**: VOTERS -- THE... New York Times

THE **1992 ELECTIONS**: PRESIDENT -- THE... New York Times

THE **1992 ELECTIONS**: THE WORLD --... New York Times

New York Times - New York Times

THE **1992 ELECTIONS**: DISAPPOINTMENT -- THE REPUBLICANS; ...

New York Times - Nov 4, 1992

By MICHAEL WINES, At a gloomy and sometimes angry post-**election ...** Mr. Bush said he had called President-elect **Bill Clinton** to congratulate him and to **...**

THE **1992 ELECTIONS** - DISAPPOINTMENT --... nytimes.com

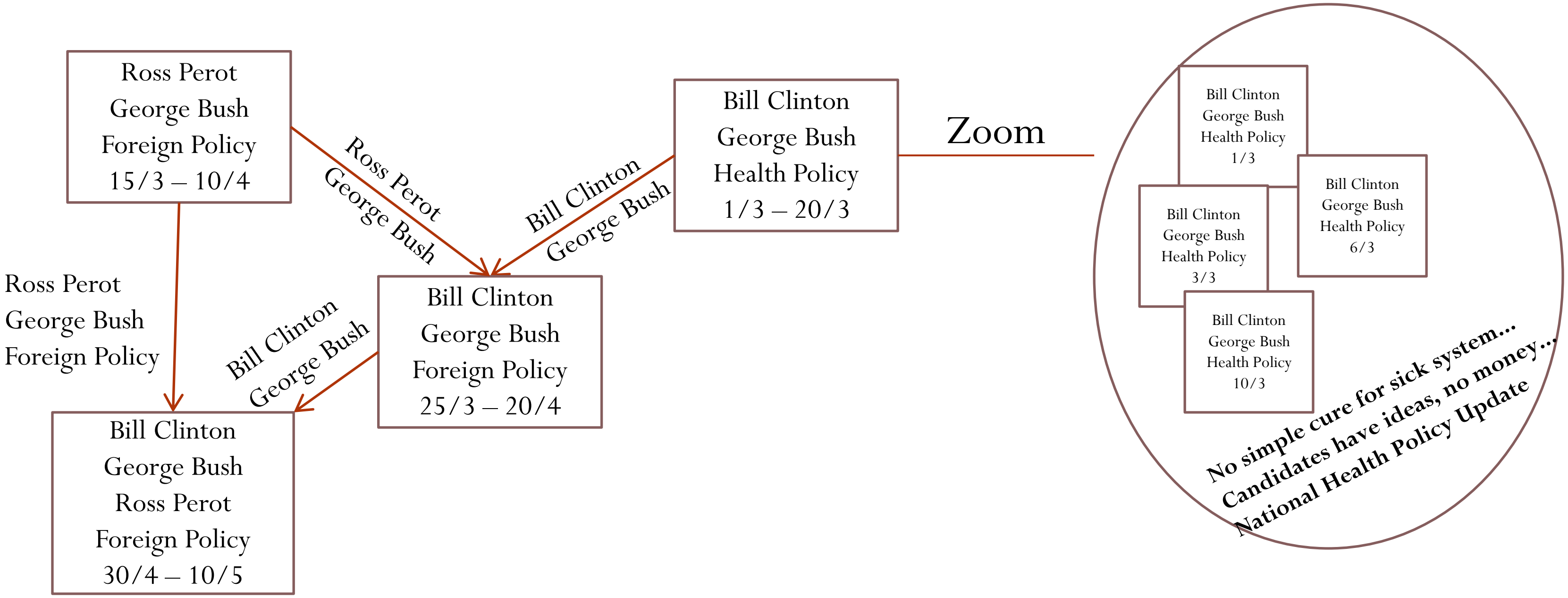
... GOP LEADER SAW MANDATE OF HIS OWN IN RESULTS OF **ELECTION**

Source: Google News

- How to frame the query?
- Suffers from lack of temporal & contexual coherence
- Impossible to get a sense of issue relevance
- Where do I **start**?

Motivation

Information Need: ***How did Bill Clinton win the 1992 elections?***



Ross Perot
George Bush
Foreign Policy
15/3 – 10/4

Ross Perot
George Bush

Bill Clinton
George Bush
Health Policy
1/3 – 20/3

Zoom

Bill Clinton
George Bush

Ross Perot
George Bush
Foreign Policy

Bill Clinton
George Bush
Foreign Policy
25/3 – 20/4

Bill Clinton
George Bush

Bill Clinton
George Bush
Ross Perot
Foreign Policy
30/4 – 10/5

Bill Clinton
George Bush
Health Policy
1/3

Bill Clinton
George Bush
Health Policy
3/3

Bill Clinton
George Bush
Health Policy
6/3

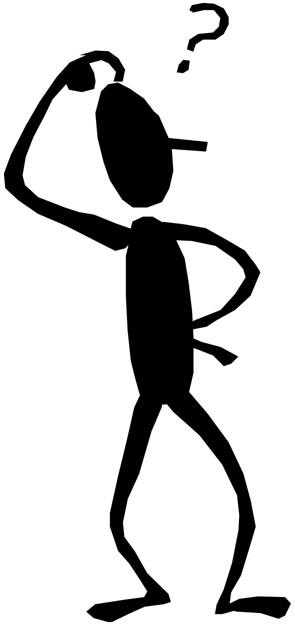
Bill Clinton
George Bush
Health Policy
10/3

No simple cure for sick system...
Candidates have ideas, no money...
National Health Policy Update

- Set all relevant actors as filters {George Bush, Ross Perot, Bill Clinton}
- Set the time period in which articles are to be studied {1992}
- Immediately reveals how their relationships evolved temporally

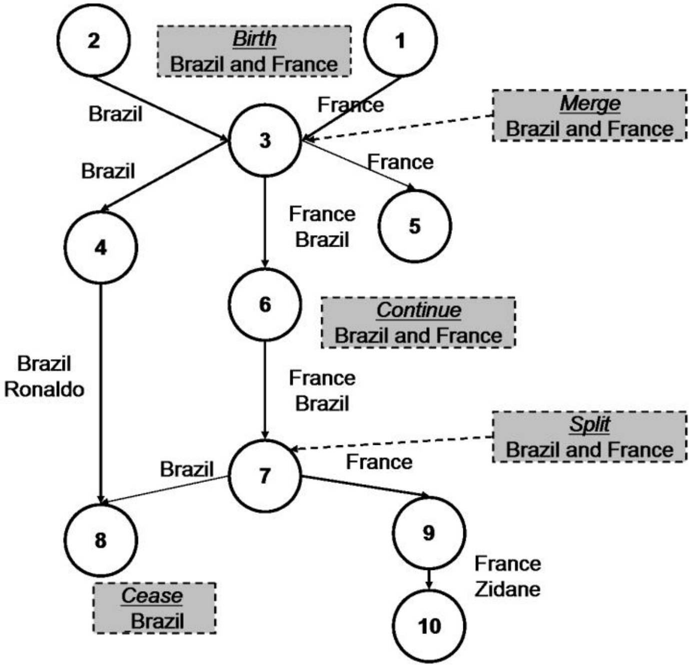
Problem Statement

- Structure news articles as a directed graph for ease of analysis
- Derive meaning/context/connections from articles
- Create an easy to use news browsing tool
- Planned as an extension of work done in [1]



A Framework for Exploration of News Corpora by Actor Evolution and Interaction [1]

- ***Actors***: Dominant entities appearing in a news article
- Place a node for each article labelled by it's actors
- Model interactions as set of 5 *transformations*



- Score each transformation and pick those above a threshold

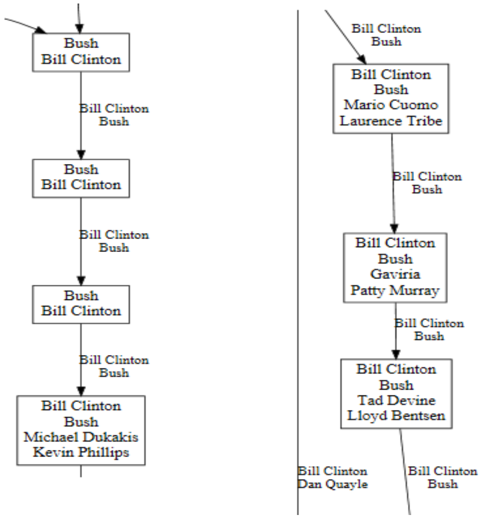
Our Dataset & Approach

- Consider the articles from NYTimes archives
- Filter out articles related to 1992 US Elections
- Extract actors for each article using Open Calais API [3]
  - returns the dominant entities, topics, social tags, relationships
- Construct the graph of these articles as discussed in [1]
  - Analyze the graph for interesting properties
  - Improve the edge coherence of graph

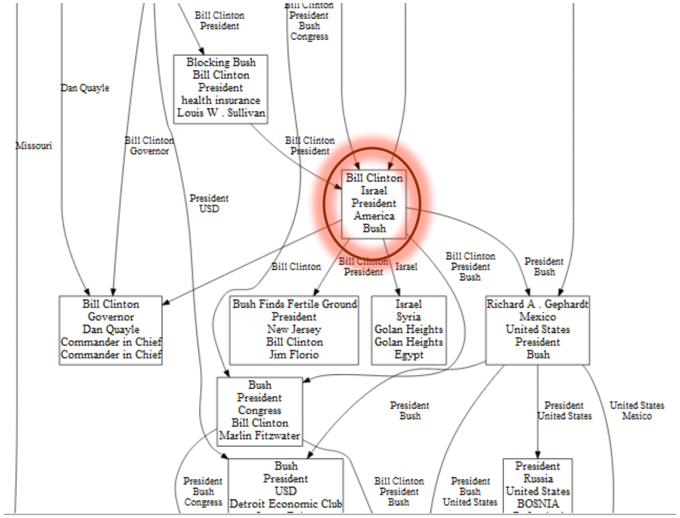
Results: Structure of the graph

- Linear chains in the graph
- Represent fast developing breaking news

- Articles of 'central' importance in the graph
- Summary articles



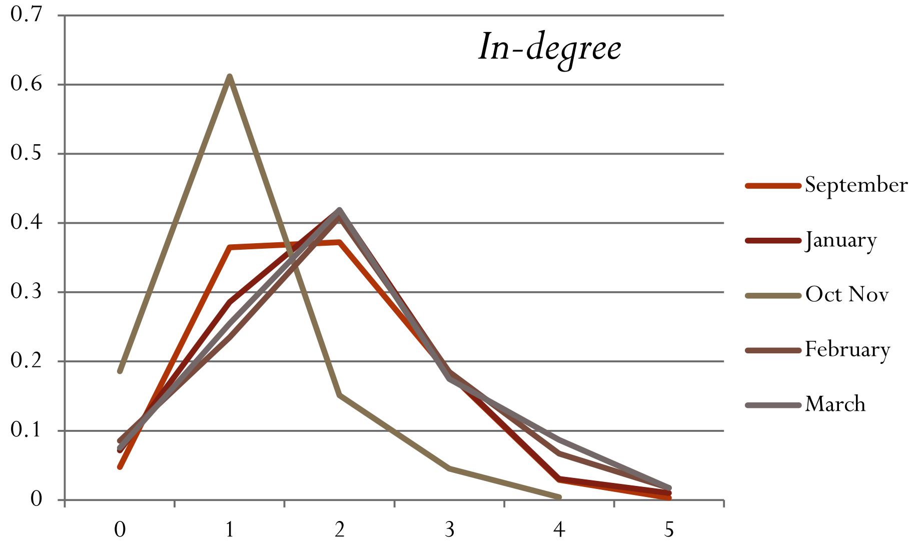
Bill Clinton & George Bush interacting with other actors



Article summarizing about all that happened the past week

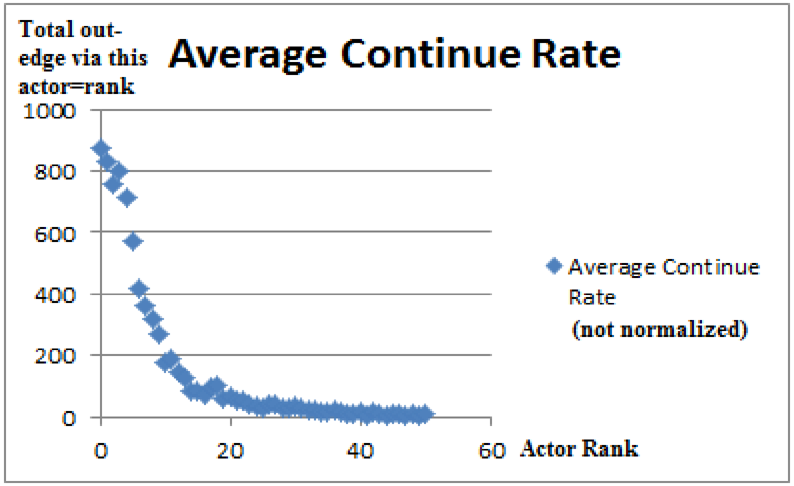
Results: Graph degree distribution

- Do degree distributions tell us something about significance of particular events and how to summarize graphs?



Results: Actor Interactions

- Unpopular actors fade out quickly
- Popular actors interact amongst each other
- The actor '**Bill Clinton**' and '**President**' are the 2 most interacting entities



Future Directions

- Current code will be rewritten so as to try other heuristics of generating graphs and summarization techniques
- Implement thematic tagging of nodes and using the tags to create independent themes from the graph
- Modeling news as an *open/closed system.*
  - Open system: Eg. Global Warming
  - Closed system: Eg. Cricket News
- Query based summarization
- Learn better summarization by user feedback

Tool Demonstration & Interface

- We can click on a node to get the corresponding article
- If we pick out a path of articles, they are connected by a coherent story
- Other visualization interfaces need to experimented

Thank You!

Questions & Comments.

References

[1] R. Choudhary, S. Mehta, A. Bagchi, and R. Balakrishnan. Towards characterization of actor evolution and interactions in news corpora. In *Advances in Information Retrieval*.

[2] Shahaf, D., Guestrin, C.: Connecting the dots between news articles. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

[3] api.opencalais.com

[4] http://en.wikipedia.org/wiki/United_States_presidential_election_1992