Bootstrapping Relation Extraction Grammars from Semantic Seeds

Hans Uszkoreit & Feiyu Xu

DFKI and
Saarland University

Thanks to Li Hong for her contributions to the experiments

Overview

- ☆ Task and motivation
- ☆ A new approach to seed-based learning for relation extraction
  - Learning extraction rules for various complexity
  - Experiments and evaluation
- ☆ Scientific questions, insights and conclusion
  - Seed-based learning in small and big worlds
  - Lessons learned and outlook

Challenge

☆ Development of a generic strategy for extracting relations/events of various complexity from large collections of open-domain free texts

Central Motivation

☆ Enable inexpensive adaptation to new relation extraction tasks/domains

Existing Unsupervised or Minimally Supervised IE Approaches

☆ **Lack of expressiveness** (Stevenson and Greenwood, 2006)

– Restricted to a certain linguistic representation, mainly verb-centered constructions e.g., subject verb object construction (Yangarber, 2003)

subject(company)-verb("appoint")-object(person)

– other linguistic constructions can not be discovered: e.g., apposition, compound NP

the 2005 Nobel Peace Prize

☆ **Lack of semantic richness** (Riloff, 1996; Agichtein and Gravano, 2000; Yangarber, 2003, Greenwood and Stevenson, 2006)

– Pattern rules cannot assign semantic roles to the arguments

subject(person)-verb("succeed")-object(person)

☆ No good method to select pattern rules, in order to deal with large number of tree patterns (Sudo et al., 2003)

☆ No systematic way to handle relations and their projections

– do not consider the linguistic interaction between relations and their projections, which is important for scalability and reusability of rules

☆ Pattern-oriented (e.g., ExDisco (Yangarber 2001))

- too closely bound to the linguistic representation of the seed, e.g.,

  subject(company) v("appoint") object(person)
- An event can be expressed by more than one pattern and by various linguistic constructions

☆ Relation and event instances as seeds (e.g., DIPRE (Brin 1998) and Snowball (Agichtein and Gravano 2000), (Xu et al. 2006))

- domain independence: it can be applied to all relation and event instances
- flexibility of the relation and event complexity: it allows n-ary relations and events
- processing independence: the seeds can lead to patterns in different processing modules, thus also supporting hybrid systems, voting approaches etc.
- Not limited to a sentence as an extraction unit

Our Approach: *DARE (1)*

☆ seed-driven and bottom-up rule learning in a bootstrapping framework

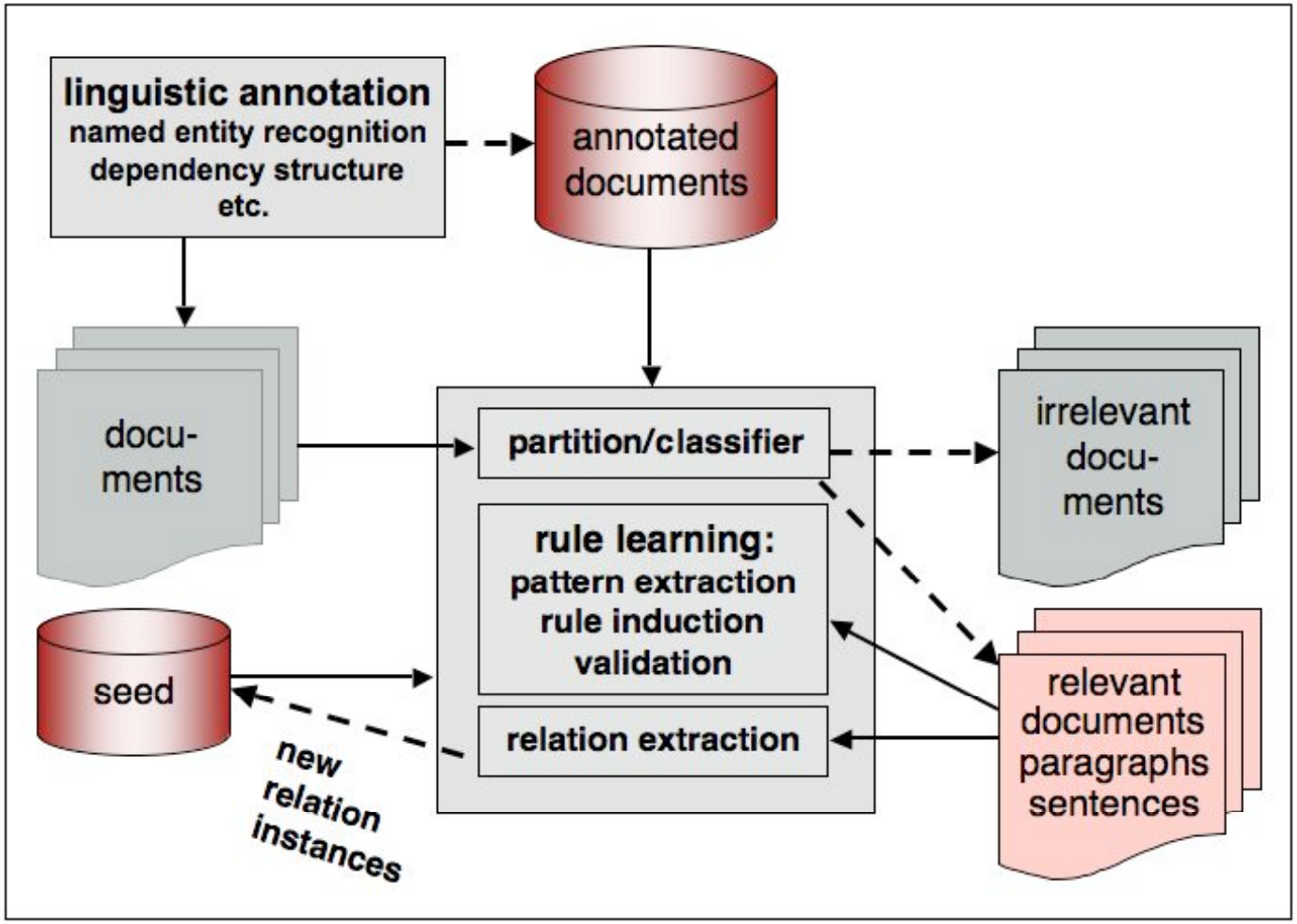
- starting from sample relation instances as seeds
  - complexity of the seed instance defines the complexity of the target relation
- pattern discovery is bottom-up and compositional, i.e., complex patterns are derived from simple patterns for relation projections
- bottom-up compression method to cluster and generalize rules
- only subtrees containing seed arguments are pattern candidates
- pattern rule ranking and filtering method considers two aspects of a pattern
  - its domain relevance and
  - the trustworthiness of its origin

Our Approach: *DARE (2)*

➡ Compositional rule representation model

- support the bottom-up rule composition
- expressive enough for the representation of rules for various complexity
- precise assignment of semantic roles to the slot arguments
- reflects the precise linguistic relationship among the relation arguments and reduces the template merging task in the later phase
- the rules for the subset of arguments (projections) may be reused for other relation extraction tasks.

DARE System Architecture



Algorithm

1. Given
   - A large corpus of un-annotated and un-classified documents
   - A trusted set of relation or event instances, initially chosen ad hoc by the user, the seed, normally, one or two.
2. NLP annotation
   - Annotate the relevant documents with named entities and dependency structures
3. Partition
   - Apply seeds to the documents and divide them into relevant and irrelevant documents
     A document is relevant, if its text fragments contain a minimal number of relation arguments of a seed
   - Paragraph/sentence retrieval
4. Rule learning
   - Extract patterns
   - Rule induction/compression
   - Rule validation
5. Apply induced rules to the same document set
6. Rank new seeds
7. Stop if no new rules and seeds can be found, else repeat 3-6

☆ Target relation

<*recipient*, *prize*, *area*, *year*>

☆ Example

Mohamed ElBaradei won the 2005 Nobel Peace Prize on Friday for his efforts to limit the spread of atomic weapons.

Example Rules

Rule name:: prize_area_year_1

Rule:: (3 year)(1 prizename) (2 areaname) 'Prize'

Output:: $\langle 1 Prize, 2 Area, 3 Year \rangle$

Rule name:: recipient_prize_area_year_1

Rule::

$$\begin{bmatrix} \text{verb} & \begin{bmatrix} \text{mode} & \text{active} \\ \text{wordform} & \text{"win"} \end{bmatrix} \\ \text{subject} & \begin{bmatrix} \text{recipient} & 1\ \text{Person} \\ \text{rule} & \text{recipient\_1:: } \langle 1 Person \rangle \end{bmatrix} \\ \text{object} & \begin{bmatrix} \text{prize} & 2\ \text{Prize} \\ \text{area} & 3\ \text{Area} \\ \text{year} & 4\ \text{Year} \\ \text{rule} & \text{prize\_area\_year\_1:: } \langle 2 Prize, 3 Area, 4 Year \rangle \end{bmatrix} \end{bmatrix}$$

Output:: $\langle 1 Recipient, 2 Prize, 3 Area, 4 Year \rangle$

Mohamed ElBaradei won the 2005 Nobel Peace Prize on Friday for his efforts to limit the spread of atomic weapons

☆ **prize_area_year_1**:

extracts a ternary projection instance <*prize, area, year*> from a noun phrase compound

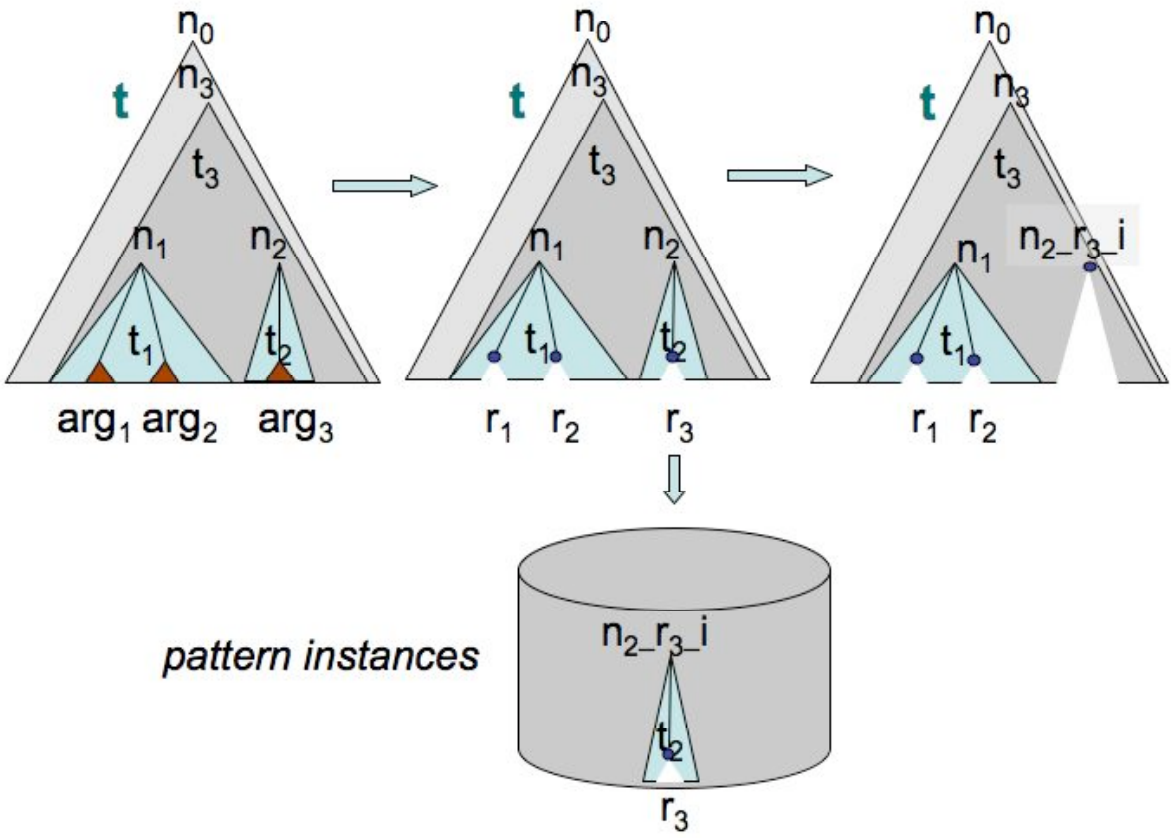
☆ **recipient_ prize_area_year_1**:

triggers **prize_area_year_1** in its object argument and extracts all four arguments.

Rule Components

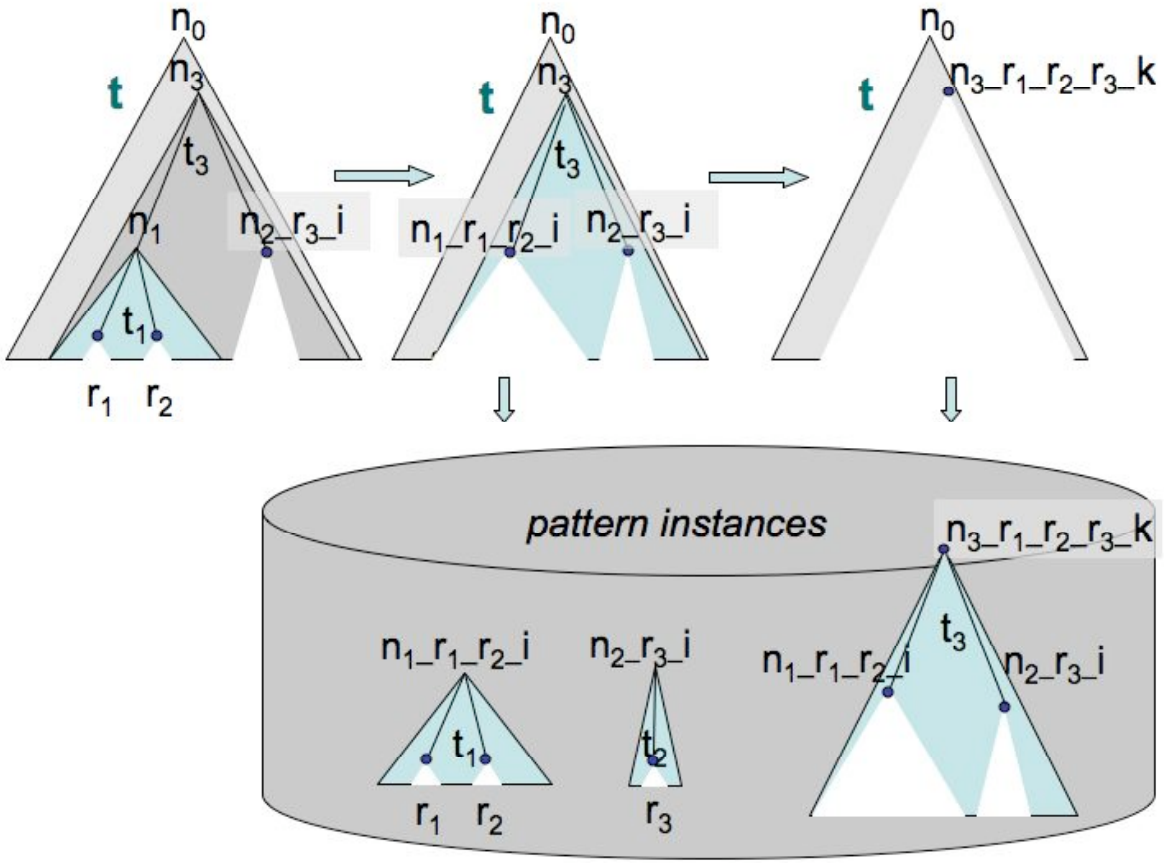
1. rule name: r_i;
2. output: a set A containing the n arguments of the n-ary relation, labelled with their argument roles;
3. rule body in AVM format containing:
   - a possibly empty set R_i of ***DARE*** rules, each of which extracts some proper subset of A;
   - a possibly empty set of constraints C_i defining which functional arguments in r_i call which rules in R_i;
   - rule-specific linguistic labels (e.g., **subject**, **object**, **head**, **mod**), derived from the linguistic analysis.

Pattern Extraction Step 1



1. replace all terminal nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;
2. identify the lowest nonterminal nodes N1 in t that dominate at most one argument (possibly among other nodes).
3. substitute N_1 by nodes labelled with the seed argument roles and their entity classes
4. prune the subtrees dominated by N_1 from t and add these subtrees into P. These subtrees are assigned the argument role information and a unique id.

Pattern Extraction Step 2



For i=2 to n

1. find the lowest nodes N_1 in t that dominate in addition to other children only i seed arguments;
2. substitute N_1 by nodes labelled with the i seed argument role combination information (e.g., r_i_r_i) and with a unique id.
3. prune the subtrees T_i dominated by N_i in t;
4. add T_i together with the argument, role combination information and the unique id to P

Event Instance as Seed

Here a relation-seed is a quadruple of 4 entity types

event &

[
Prize Name : *prize_name*
Prize Area : *area_name*
Recipient List : list of *person*
Year: *year*
]

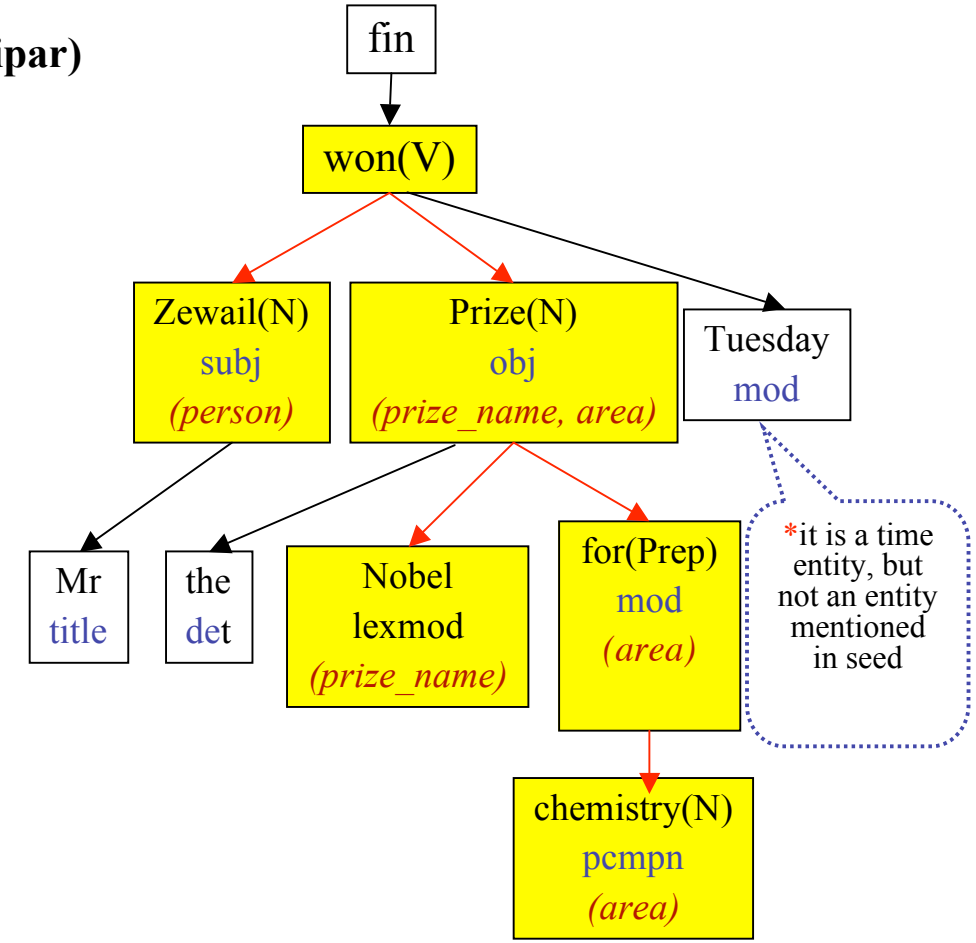
Examples in xml

```
<seed id="1">
   <prize name="Nobel"/>
   <year>1999</year>
   <area name="chemistry"/>
   <recipient>
        <person>
             <name>Ahmed H. Zewail</name>
             <surname>Zewail</surname>
             <gname>Ahmed</gname>
             <gname>H</gname>
        </person>
   </recipient>
</seed>
```

Seed: *(Nobel, chemistry, [Ahmed H. Zewail], 1999)*

Sentence: ***Mr. Zewail won the Nobel Prize for chemistry Tuesday.***

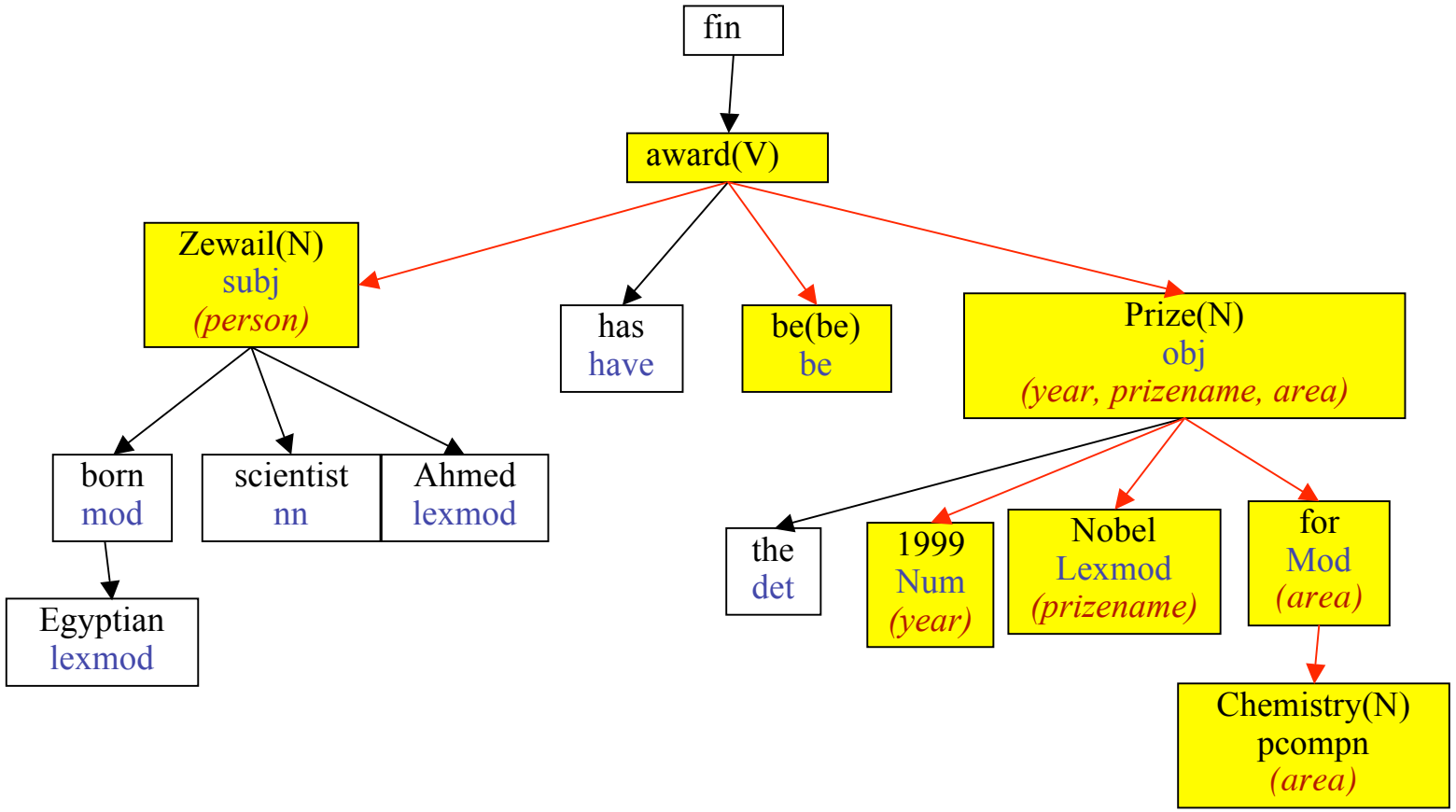
Parse Tree (SProUT + Minipar)



Example 2

Seed: ***(Nobel, chemistry, [Ahmed H. Zewail], 1999)***

Sentence: ***Egyptian-born scientist Ahmed Zewail has been awarded the 1999 Nobel Prize for Chemistry.***



☆ Which kind of sentences could represent an event?

complexity	matched sentence	event sentence	Relevant sentences in %
4-ary	36	34	94.0
3-ary	110	96	87.0
2-ary	495	18	3.6

Table 1. distribution of the seed complexity

Distribution of Relation Projections

combination (3-ary, 2-ary)	matched sentence	event sentence	relevant sentences in %
person, prize, area	103	91	82%
person, prize, time	0	0	0%
person, area, year	1	1	100%
prize, area, year	6	4	68%
person, prize	40	15	37%
person, area	123	0	0%
person, year	8	3	37%
prize, area	286	0	0%
prize, year	25	0	0%
area, year	12	0	0%

Table 2. distribution of entity combinations

- ☆ Two domains
  - Nobel Prize award: *<recipient, prize, area, year>*
  - management succession: *<Person_In, Person_Out, Position, Organisation>*

- ☆ Test data sets

Data Set Name	Files	Volume
Nobel Prize A (1999-2005)	2296	12,6 MB
Nobel Prize B (1981-1998)	1032	5,8 MB
MUC-6	199	1 MB

☆ Conditions and Problems

- Complete list of Nobel Prize award events from online portal Nobel-e-Museum
- No gold-standard evaluation corpus available

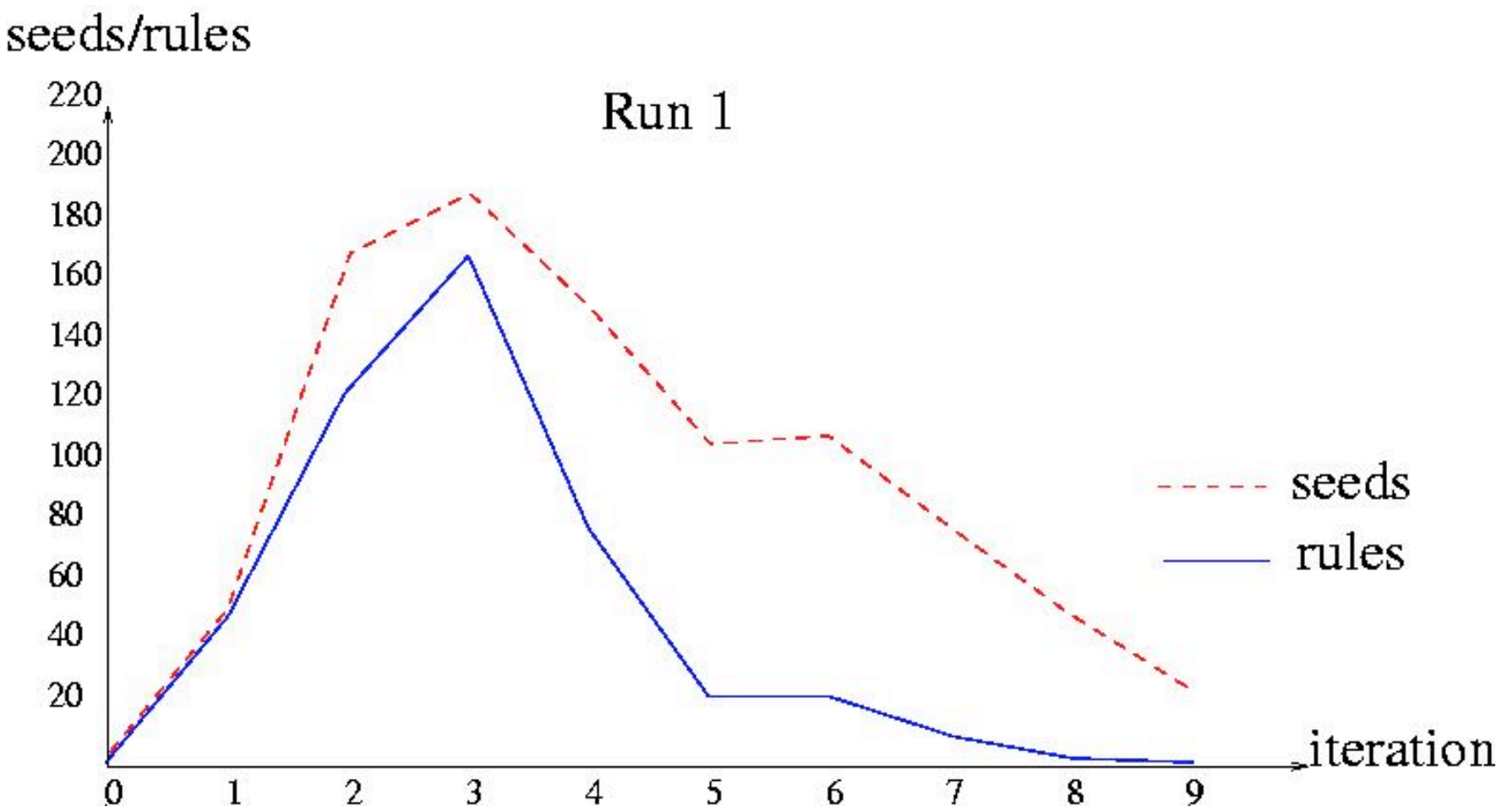
☆ Solution

- our system is successful if we capture one instance of the relation tuple or its projections, namely, one mentioning of a Nobel Prize award event. (Agichtein and Gravano, 2000)
- construction of so-called *Ideal* tables that reflexe an approximation of the maximal detectable relation instances
  - The Ideal tables contain all Nobel Prize winners that co-occur with the word “Nobel” in the test corpus and integrate the additional information from the Nobel-e-Museum

Evaluation Against Ideal Tables

Data Set	Seed	Precision	Recall
Nobel Prize A	<[Zewail, Ahmed H], nobel, chemistry, 1999>	**71.6%**	**50.7%**
Nobel Prize B	<[Sen, Amartya], nobel, economics, 1998>	**87.3%**	**31.0%**
Nobel Prize B	<[Arias, Oscar], nobel, peace, 1987>	**83.8%**	**32.0%**

Iteration Behavior (Seed vs. Rule)



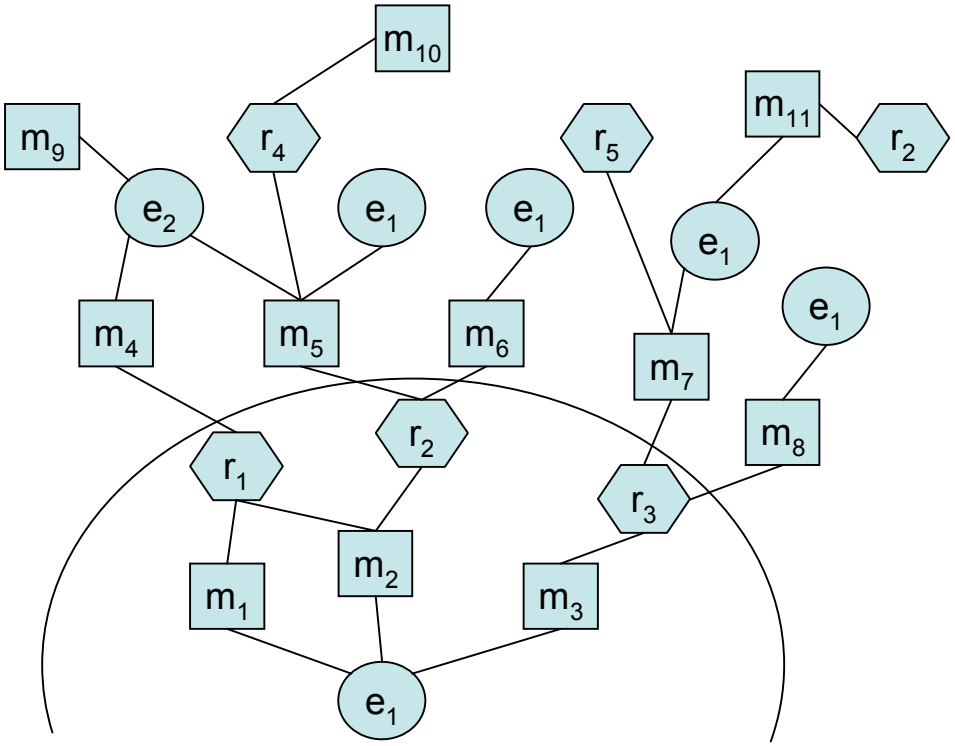
The Dream

- ☆ Wouldn‘t it be wonderful if we could always automatically learn most or all relevant patterns of some relation from one single semantic instance!
- ☆ Or at least find all event instances. (IDEAL Tables or Completeness)
- ☆ This sounds too good to be true!

☆ As scientists we want to know:

- Why does it work for some tasks?
- Why doesn‘t it work for all tasks?
- How can we estimate the suitability of domains?
- How can we deal with less suitable domains?

Start of Bootstrapping (simplified)



Can we reach all events in the graph?

By how many steps?

From any event instance?

Abstraction

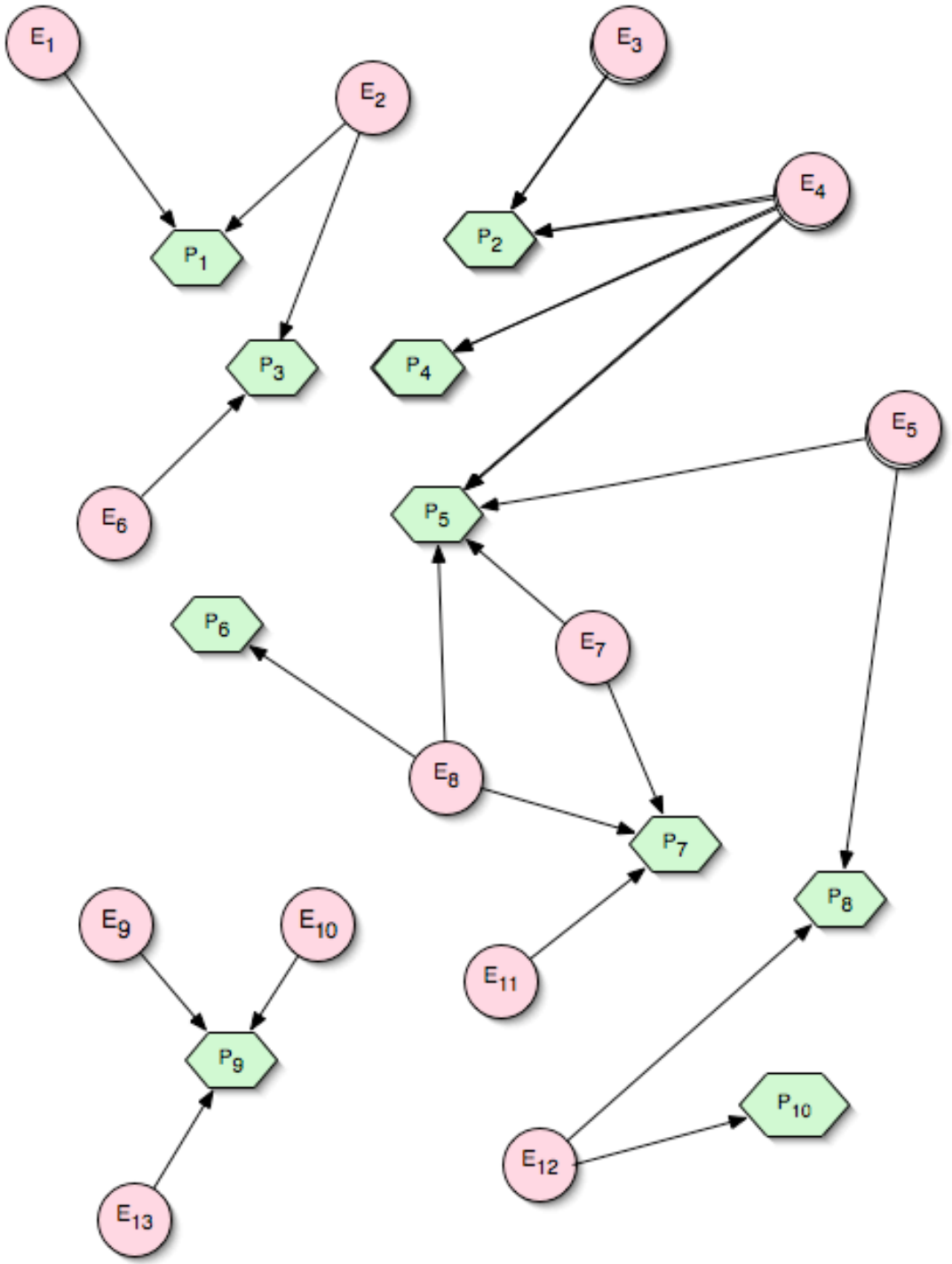
bipartite graph

two types of vertices

E_i = *event instance*
P_j = *linguistic pattern*

relevant properties:

- ☆ two degree distributions
- ☆ connectedness
- ☆ average and maximum path lengths between events



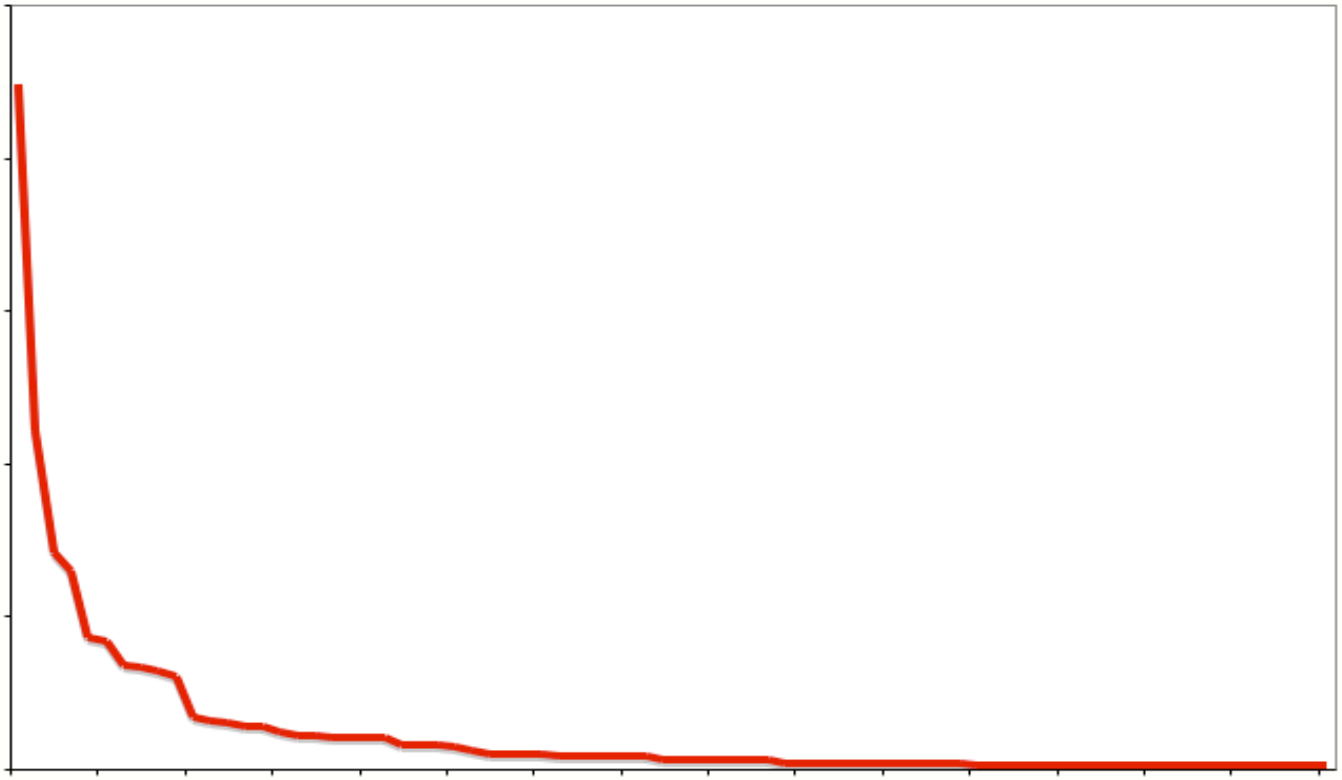
Two Distributions

1. Distributions of Pattern in Texts

2. Distribution of Mentionings to Relation Instances

Two Distributions

General distribution of patterns in texts probably follows Church's Conjecture: Zipf distribution (a heavy-tailed skewed distribution)



Distribution of Mentionings to Events

- ☆ Distribution of mentionings to relation instances (events) differs from one task to the other.
- ☆ The distribution reflects the redundancy in textual coverage of events.
- ☆ Distribution depends on text selection, e.g. number of sources (newspapers, authors, time period)

  example 1: several periodicals report on Nobel Prize events

  example 2: one periodical reports on management succession events

☆ Degree Distribution gives the probability distribution of degrees in a complex network

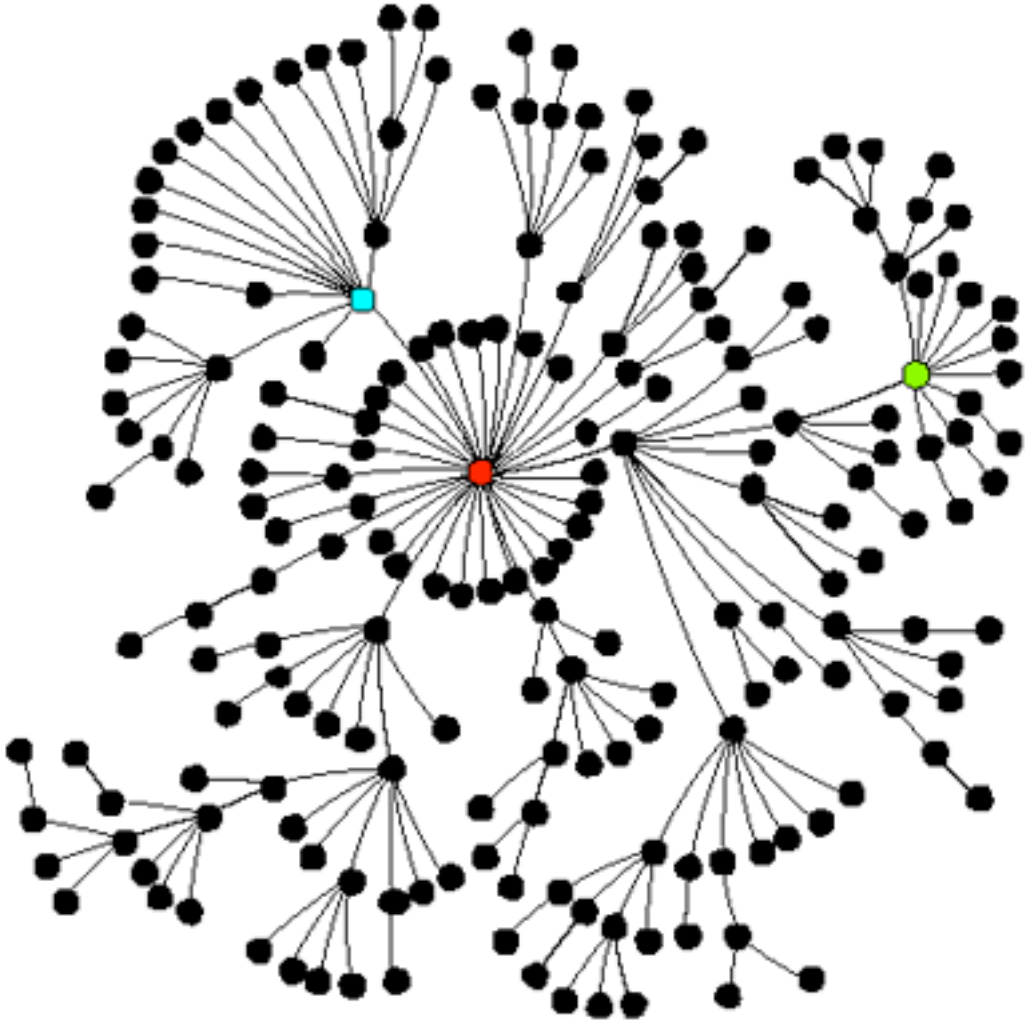
$$p(k) = \sum_{v \in V \mid \deg(v)=k} 1$$

☆ scale-free networks

$$\mathbf{P(k)} \sim \mathbf{k}^{-\gamma}$$

Zipf-like distribution (heavy-tailed skewed distribution) of degrees

Example of Scale-Free Nets



Small-World Property

Networks exhibiting the small-world property

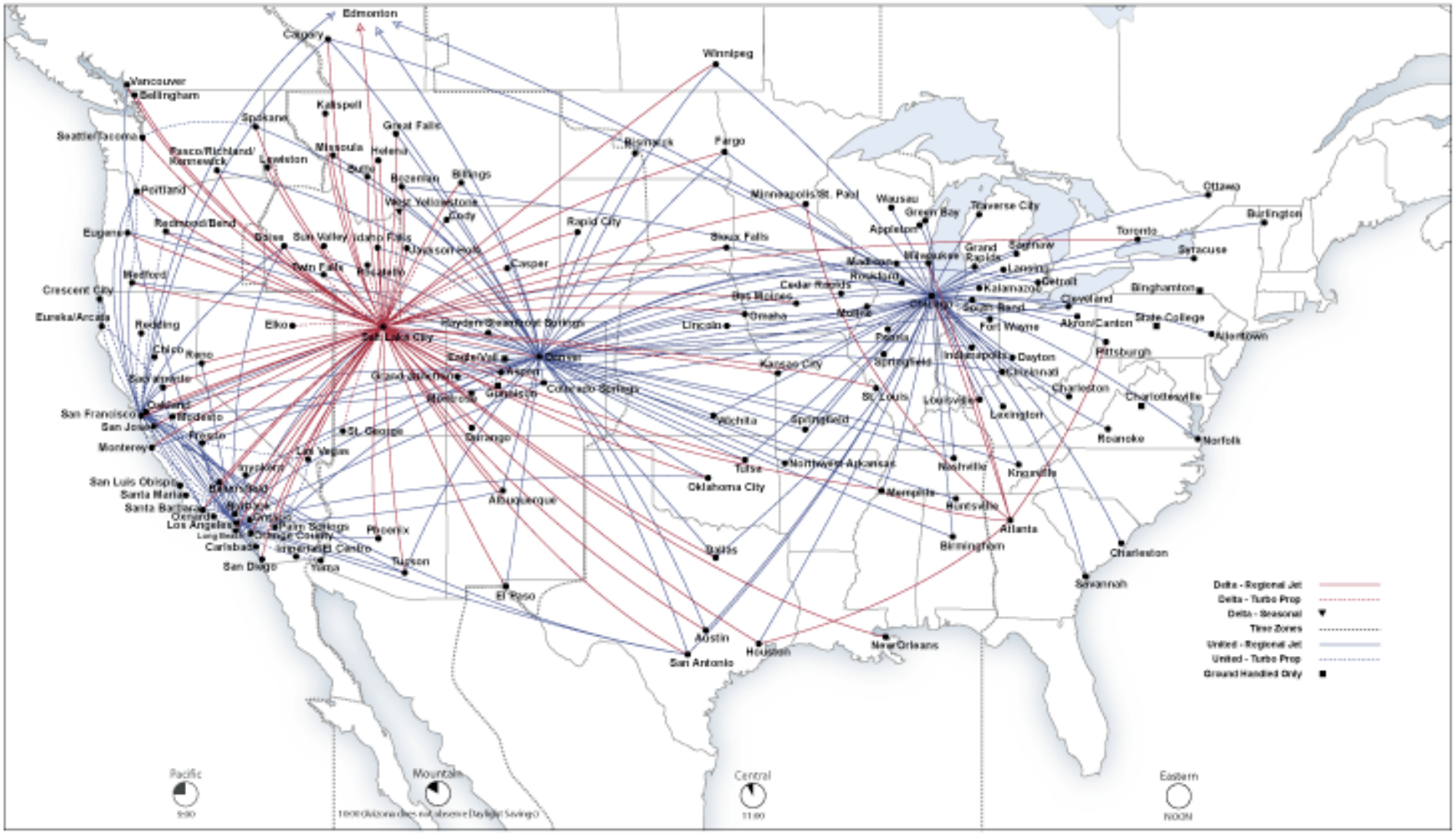
- social networks (max path-length 5-7)
- co-authorship networks (Erdös number)
- Internet
- WWW
- air traffic route maps (max. 3 hops)

Networks that do not exhibit the small-world property

- road networks
- railway networks
- kinship networks

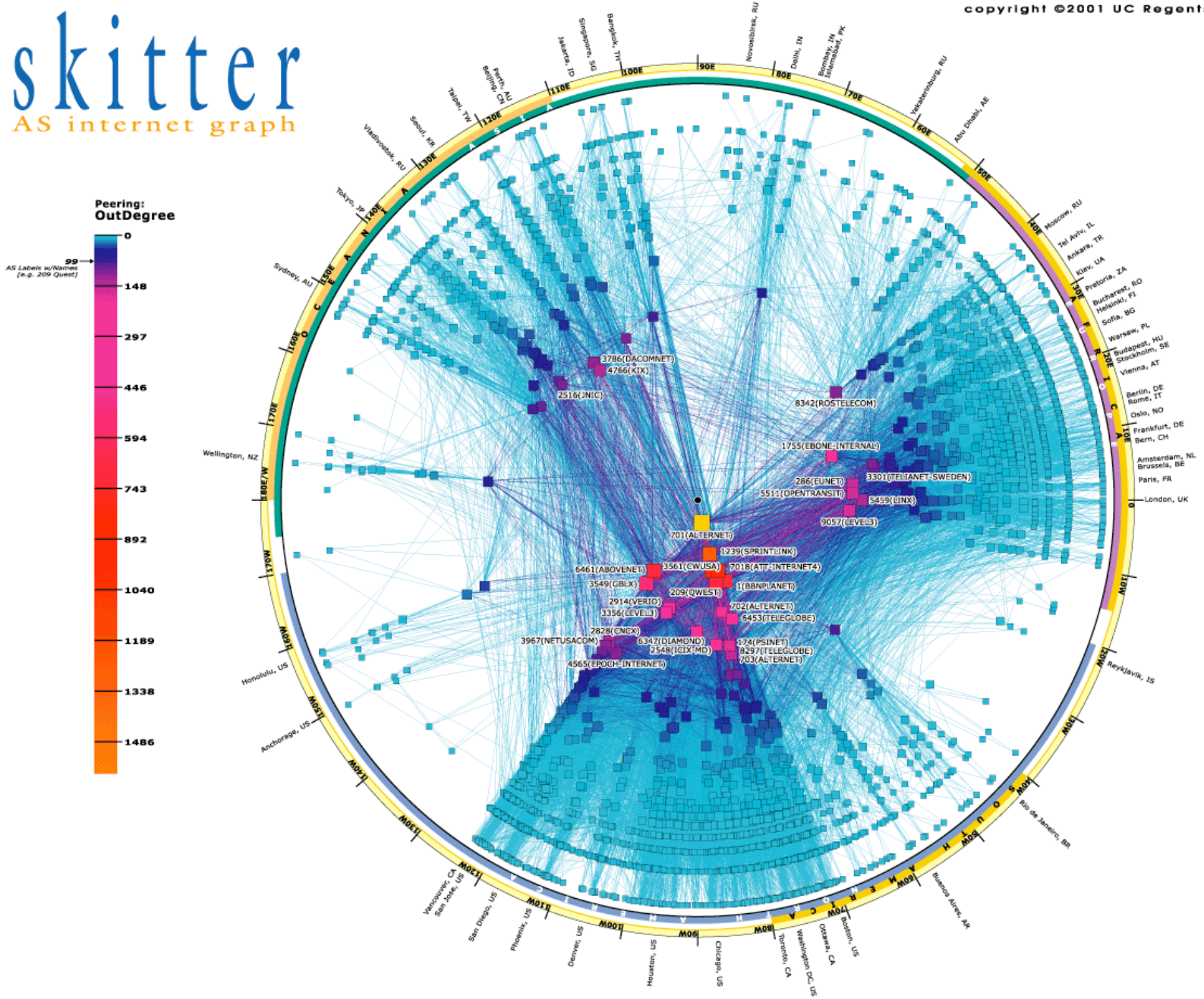
Airline Route Networks

HANS USZKOREIT & FEIYU XU 07



skitter

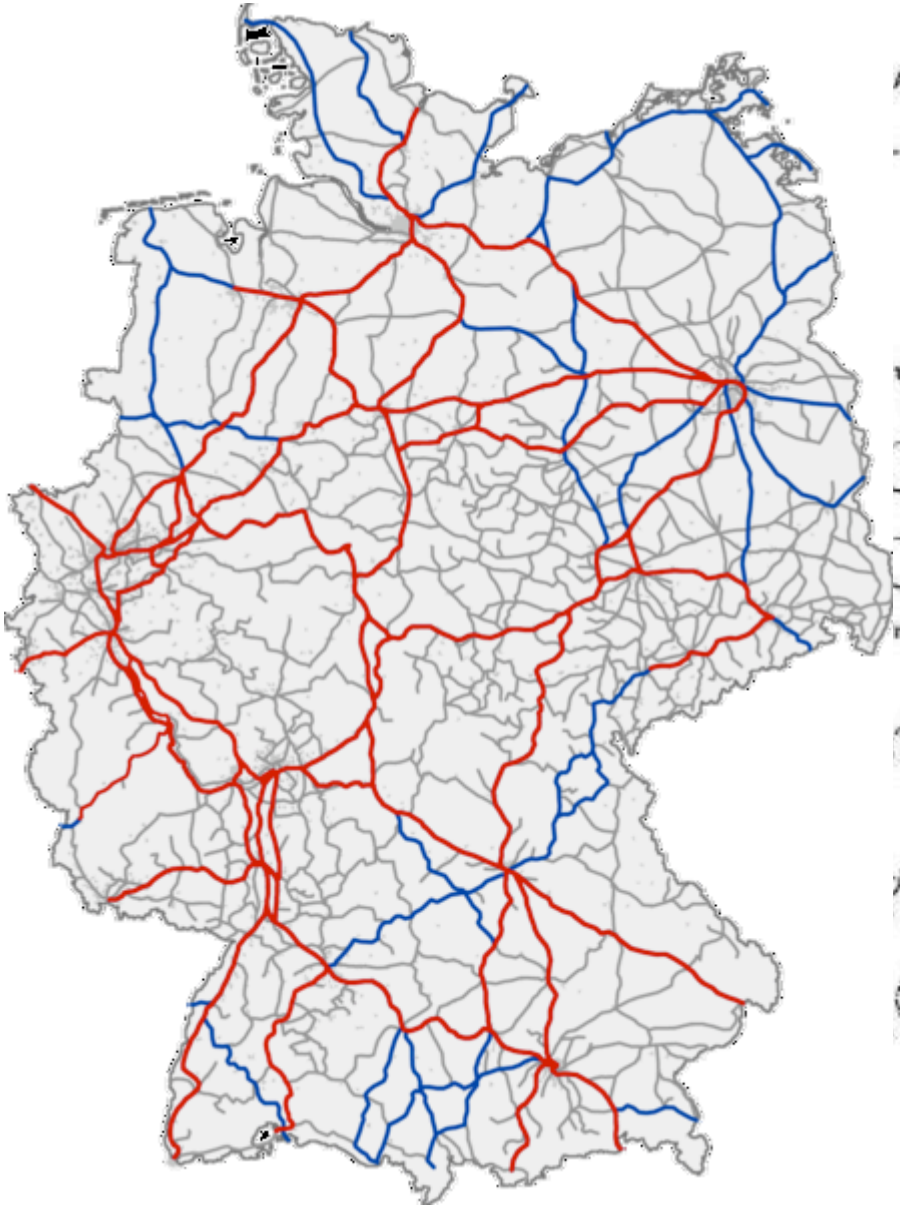
AS internet graph

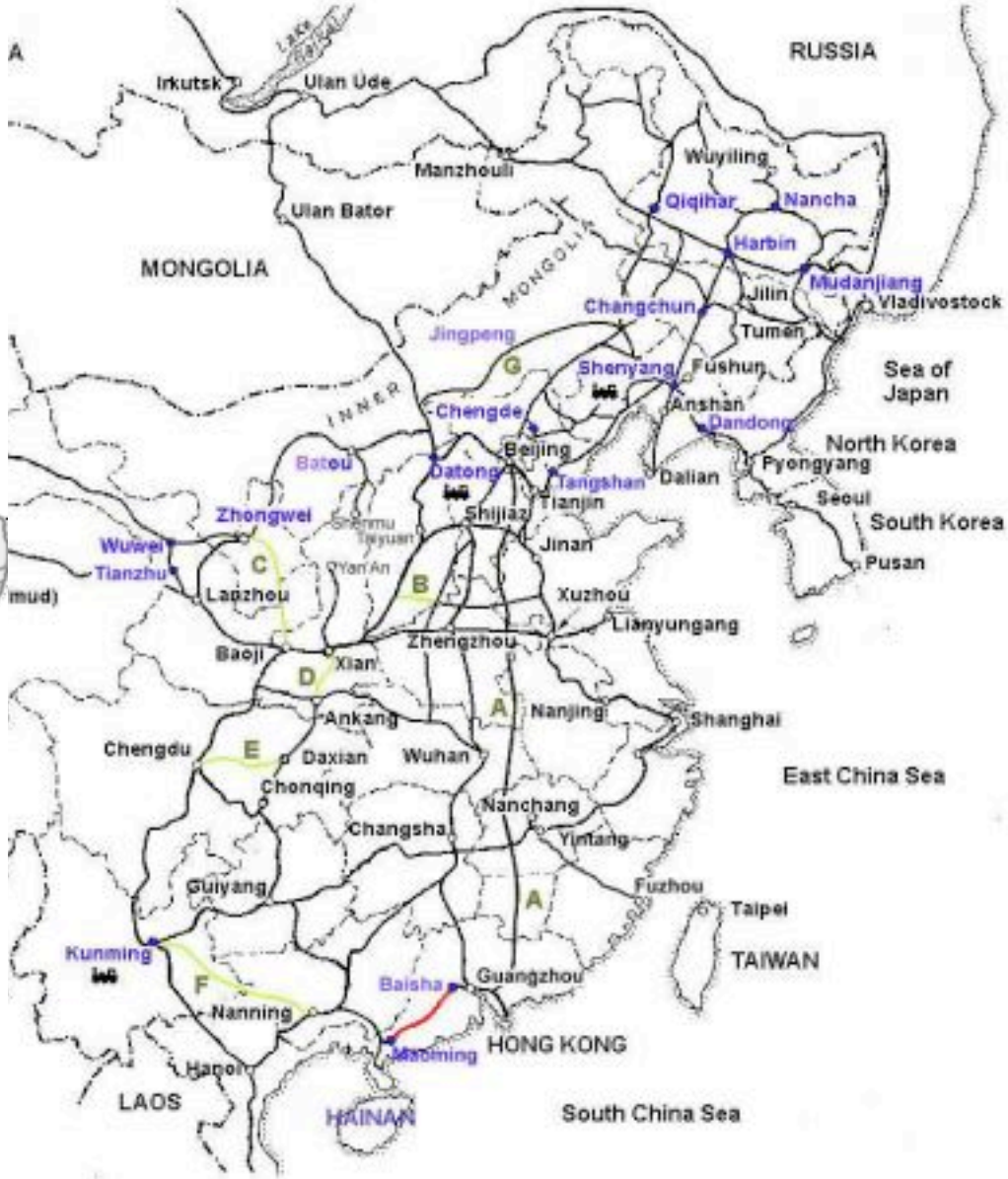


cooperative association for internet data analysis ○ san diego supercomputer center ○ university of california, san diego
9500 gilman drive, mc0505 ○ la jolla, ca 92093-0505 ○ tel. 858-534-5000 ○ http://www.caida.org/

caida
www.caida.org

CAIDA is a program of the University of California's San Diego Supercomputer Center (UCSD/SDSC)
skitter is supported by DARPA NGI Cooperative Agreement N66001-98-2-8922, NSF ANIR Grant NCR-9711092 and CAIDA members




RUSSIA
Irkutsk
Ulan Ude
Manzhouli
Wuyiling
Qiqihar
Nancha
Ulan Bator
MONGOLIA
Harbin
Mudanjiang
Jilin
Vladivostock
Changchun
Tumen
Jingpeng
Shenyang
Fushun
Sea of Japan
Anshan
Dandong
Chengde
Beijing
North Korea
Pyongyang
Datong
Tangshan
Dalian
Seoul
Tianjin
Zhongwei
Shijiaz
South Korea
Jinan
Wuwei
Tianzhu
Lanzhou
Xuzhou
Lianyungang
Zhengzhou
Baoji
Xian
Ankang
Nanjing
Shanghai
Chengdu
Daxian
Wuhan
East China Sea
Chongqing
Nanchang
Changsha
Yintang
Guiyang
Fuzhou
Taipei
TAIWAN
Kunming
Baisha
Guangzhou
Nanning
HONG KONG
Maoming
Hanoi
LAOS
HAINAN
South China Sea

- ☆ If both distributions follow a skewed distribution and if the distributions are independent from each other, then we get a scale-free network in the broader sense of the term.

- ☆ For each type of vertices we get strong hubs. This leads to very short paths (for most connections).

- ☆ However, there are degrees of the small-world property.
- ☆ Small World Networks are further optimized if there are forces beyond probability that cause hubs to be directly connected.

☆ Enlarging the domain

Pulitzer Prize --> all Prizes

☆ selecting Carrier Domains (parallel learning domains)

Pulitzer Prize --> Nobel Prize
Ernst Winter Preis --> Nobel Prize
Fritz Winter Preis --> Nobel Prize

Other Discovered Award Events

Academy Award
actor % (Cannes Film Festival's Best Actor award)
American Library Association Caldecott Award
American Society
award
Blitzker
Emmy
feature % (feature photography award)
first % (the first Caldecott Medal)
Francesca Primus Prize
gold % (gold medal)
Livingston Award
National Book Award
Newbery Medal
Oscar
P.G.A
PEN/Faulkner Award
prize
reporting % (the investigative reporting award)
Tony
Tony Award
U.S. Open

But also:

nomination
$1 million
$29,000
about $226,000
praise
acclaim
discovery
doctorate
election

☆ enlarging the text base for finding seeds and patterns

- New York Times MUC data --> general press corpora
- New York Times MUC data --> WWW

☆ enlarging the text base for finding new seeds

- New York Times MUC data --> WWW
- German Press Data --> English Press Data

Summary

☆ Our approach works with semantic seeds.

☆ It learns rules for an n-ary relation and its projections.

☆ Rules mark the slot-filler with their roles.

☆ For some relation extraction tasks, the semantic seed based bootstrapping approach works surprisingly well.

☆ For others, it still works to some degree.

☆ Our deeper understanding of the problem helps us to select or prepare data for effective learning.

Next Steps

- ☆ Go beyond the sentence.
- ☆ Investigate properties of relations w.r.t. data.
- ☆ Try to describe them as graph properties.
- ☆ Try out auxiliary data sets (such as the Web).
- ☆ Try out deep processing: extract patterns from RMRS with extended ERG (first tests by Zhang Yi 80% coverage for Nobel prize sentences, 61% for management succession)

The material presented here has been submitted for publication.
An earlier stage of the results was published in:

Feiyu Xu, Hans Uszkoreit & Hong Li: Automatic Event and Relation Detection with Seeds of Varying Complexity. In Proceedings of the AAAI 2006 Workshop Event Extraction and Synthesis, Boston, July, 2006.