An Introduction to the World of Small Worlds

J. Ramanand

March 12, 2007

Meet John Doe

- How many links connect you to the Director of IIT Bombay?

Meet John Doe

- How many links connect you to the Director of IIT Bombay?
- How many links connect you to the President of India?

Meet John Doe

- How many links connect you to the Director of IIT Bombay?
- How many links connect you to the President of India?
- How many links connect you to Adam Gilchrist/Nelson Mandela/Tim Berners-Lee?

Meet John Doe

- How many links connect you to the Director of IIT Bombay?
- How many links connect you to the President of India?
- How many links connect you to Adam Gilchrist/Nelson Mandela/Tim Berners-Lee?
- Sociological studies show that the diameter of human social graph is less than 10. *(*very, very small!)
- The concept of **"Six Degrees of Separation"**

Graphs and Measures

- Measures
  - Average Shortest Path Length
  - Clustering Coefficient
  - Degree Distribution
- Random Graphs: low Avg. SP
- Regular Graphs: high Avg. SP
- Small World Graphs: low Avg. SP

Cluster Coefficient

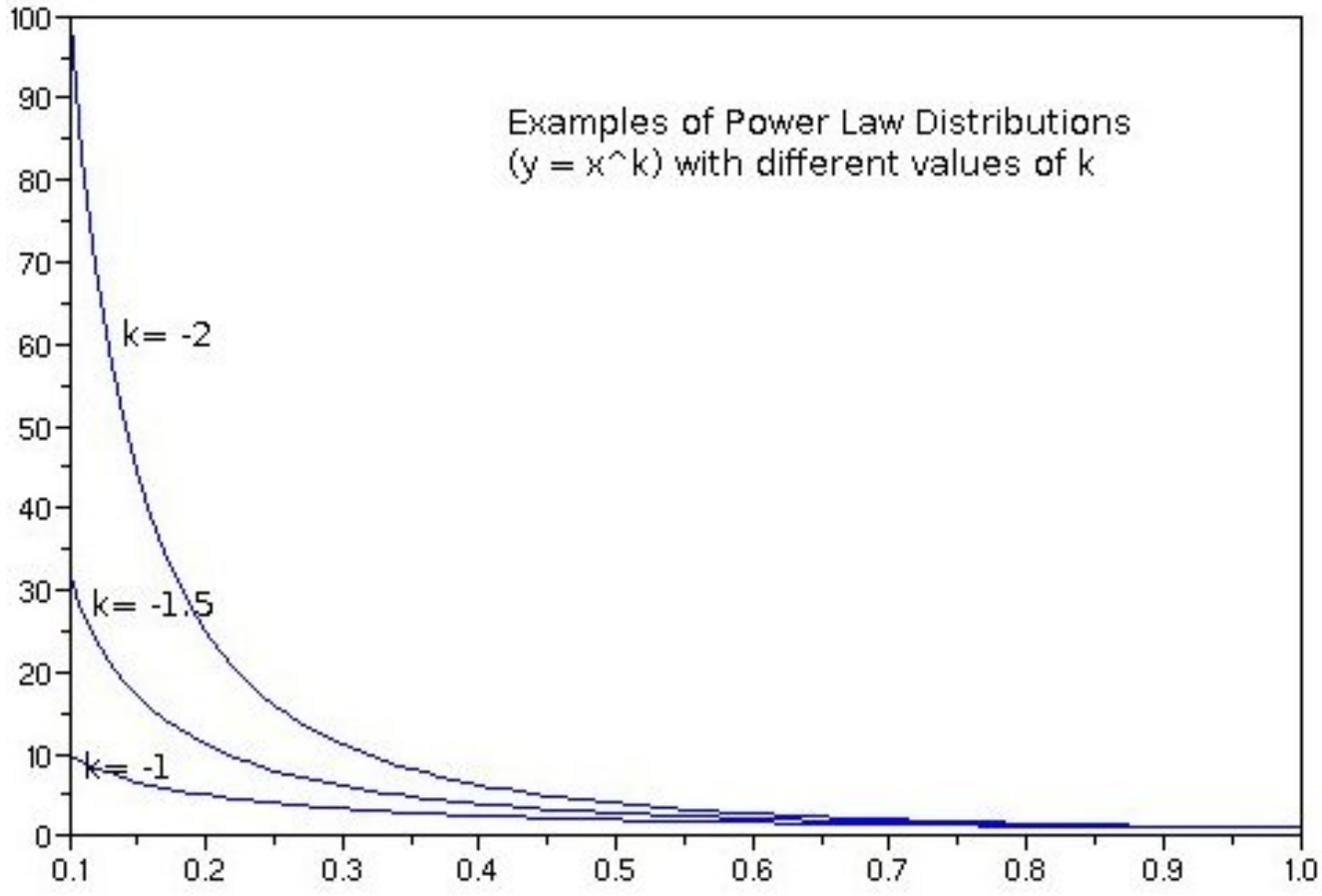
- Measures what fraction of neighbours of a node are related to each other
- *Cluster Coefficient* C_i for a node i (with degree k_i) of a directed graph:

$$C_i = \frac{|E(\Gamma_i)|}{2 \times \binom{k_i}{2}}$$

  where Γ_i is the subgraph made of i and its neighbours, $|E(\Gamma_i)|$ is the number of edges of the subgraph, and $2 \times \binom{k_i}{2}$ is the total number of possible edges in Γ_i.
- Random Graphs: low CC ($CC << 1$)
- Regular Graphs: high CC ($0.4 <= CC <= 0.7$)
- Small World Graphs: high CC ($0.4 <= CC <= 0.7$)

Degree Distribution

- Degree Distribution ($P(k)$): for each degree k, the proportion of total number of nodes that have degree k
- A power-law shape is observed; *scale-free* nature of graphs
- $y = x^k$



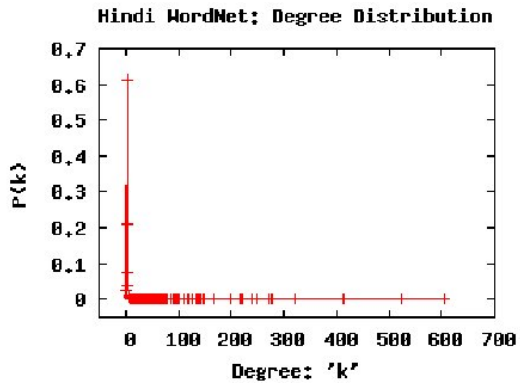
So why do we study *Small Worlds*?

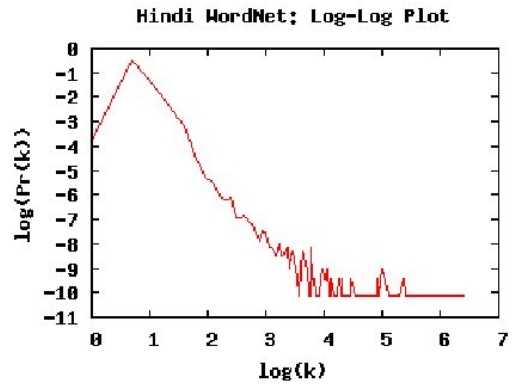
- Because several complex networks seem to exhibit this
- Examples: The WWW graph; Movie-costarring graphs; Academic citations; Biological oscillators; Social networks
- (Recommended Exercise: Play the *Kevin Bacon Game*! (http://oracleofbacon.org/))
- Learn how such kind of information is organised
- A few important Hubs; the majority are on the periphery; presence of Long tails;
- Robustness of such graphs

But why study *Small Worlds* in this class?

- Because we also see these properties in language organisation
- More precisely, we see them in word/concept knowledge bases such as (good old) Wordnet
- A Wordnet can be thought of as a directed graph where
  nodes == synsets (concepts)
  edges == relations (such as hypernyms, hyponyms)
- In fact, we've observed these properties even in non-English Wordnets (Hindi and Marathi)

Degree Distribution in Wordnets





Wordnet	Exponent(γ)
English WN (Nouns)	-2.063
Hindi WN	-2.592
Marathi WN	-2.841

Table: Exponents for the Degree Distributions

Some examples:

- EWN (Nouns): (city,metropolis,urban_center): 664, (law,jurisprudence): 611, (person,individual,someone,somebody,mortal,soul): 400
- EWN (Verbs): (change,alter,modify): 397, (change): 188, (be): 132
- HWN: (vyaktii, maanas, shaks, shakhs, ba.ndaa (person)) 607, (karm, karanii, kaam, kaarya, krtya, kaarvaaii, kaarvaahii (action)): 524, (avasthaa, dashaa, haalaat, sthithii, vrttii, suurat, haal, gatii (state)): 414
- MWN: (vyaktii, maaNus, isama, manushya, paTThaa, paThyaa (person)): 626, (karm, krtii, kriyaa, kaam, kaarya, krtya (action)): 546, (avasthaa, sthithii, dashaa, gat (state)): 428

Cluster Coefficient in Wordnets

Wordnet	Avg. Cluster Coefficient
English WN (Nouns)	0.526
Hindi WN	0.268
Marathi WN	0.358

Synset	Degree	CC
Hamas, Islamic_Resistance_Movement	3	0.667
air_unit	10	0.044
thing	22	0.082
cell	36	0.007
New_Testament	51	0.012
England	85	0.004
baseball	98	0.002
animal_order	102	0.010
military, armed_forces, armed_services, ...	224	0.002
law, jurisprudence	626	0.000

Average Shortest Path Length in Wordnets

Wordnet	**Average Shortest Path**	**Maximum Shortest Path**
EWN (Nouns)	8.878	20
HWN	4.378	15
MWN	4.255	20

Applications

- The conjecture is: perhaps this explains how humans store words and concepts?
- Understanding how language evolved/evolves
- Understanding how humans build up their vocabulary
- Studies of language disorders - aphasia, agrammatism
- **What we are doing:** trying to see if we can evaluate wordnets using this information

References

- A good introduction: Xiao Fan Wang, Guanrong Chen. *Complex Networks: Small-World, Scale-Free and Beyond*. IEEE Circuits and Systems Magazine, First Quarter, 2003.
- The book: Duncan Watts. *Six Degrees: The Science of a Connected Age*. Norton, New York, 2003.
- A more mathematical treatment: Duncan Watts. *The Dynamics of Networks between Order and Randomness*. Princeton University Press, 2006.
- Watts' "Small World Project" (http://smallworld.columbia.edu)
- Ricard V. Sole. *Language Networks: their structure, function and evolution*.
- Mariano Sigman, Guillermo A. Cecchi. *Global organization of the Wordnet lexicon*. Proceedings of the National Academy of Sciences of the USA. Vol. 99. Feb 2002.
- J.Ramanand, Akshay Ukey, Brahm Kiran Singh and Pushpak Bhattacharyya, *Mapping and Structural Analysis of Multilingual Wordnets*, IEEE Data Engineering Bulletin, 30(1), March 2007 (to appear :-)).