Ideas Map: mapping the relationship between ideas in publications

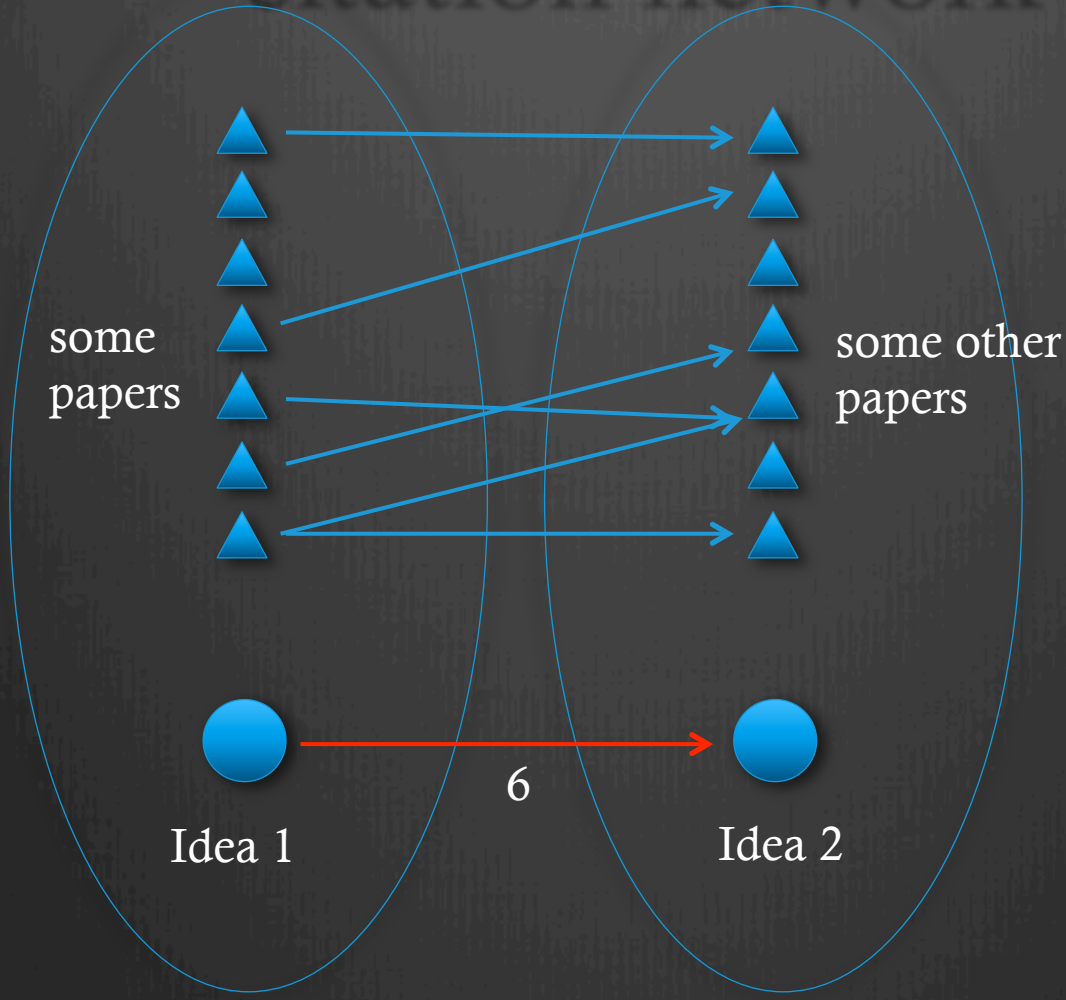
Koon-Kiu Yan, Nets

3/15/2011

Evolution of knowledge

- Ideas/concepts/scientific information pass from publications to publications
- Richard Dawkins: meme as analogous to gene
- Imagine ideas or concepts etc form a universe of knowledge, the interdependence between ideas in the knowledge universe is a network called ideas map
- A publication can be modeled as a collection of ideas
  - A novel idea appears
  - Bring two existing ideas together, but novel connection
  - Reinforce the connections between two ideas, weight
  - Example: 2008 Nobel Prize in chemistry

Construction of Ideas Map from citation network



Practically,

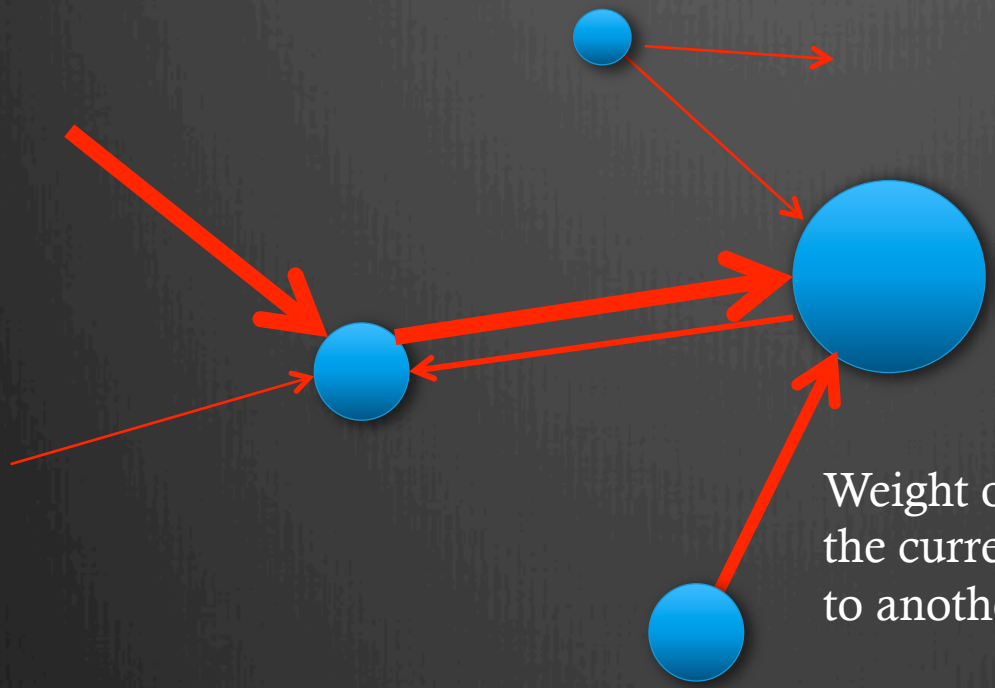
- Ideas/concepts are represented by a set of "keywords". We have explored two possibilities
- NB's keyword list, 2200 keywords
  - Scan the titles and abstracts of 6 million PubMed articles, which are interconnected by a citation network (30 million edges)
- MeSH terms, 26000 terms
  - MeSH terms are assigned to every paper in PubMed

Network centrality

Use PageRank to quantify centrality, the most important idea in the universe…

MeSH terms	NB's list
Humans	Gene
Animals	Protein
Male	Human
Female	Expression
Adult	Response
Mice	Patients
Molecular Sequence Data	Sequence
Middle Aged	Disease
Amino acid Sequence	Mechanism
Rats	Receptor
Aged	Treatment
Cell line	Model
Mutation	Binding
Kinetics	icu
Adolescent	Development
Cells, Cultured	Growth
RNA, messenger	Evidence
Time factors	Virus
E. coli	race

Visualization of an idea map



Size of node proportional to the PageRank

Weight of edge proportional to the current flow from one node to another

- Importance of an idea
- Importance of idea A to idea B: 1. adjacent nodes 2. any two nodes

What can be done?

- Alternate way of construction: instead of direct counting the number of citations between two keywords, we could introduce some statistical measures
- Bottleneck: the choice of keywords
  - NB's list
  - MeSH terms
  - Wikipedia
  - medical dictionary
  - Topic Modeling

I haven't touched evolution yet.

- Many questions can be asked based on the ideas map. Inspiration from the book "The Structure of scientific revolutions" by Thomas Kuhn
  - How to define a scientific paradigm?
  - Tracking the gain in edge weight. Distinguish "normal science" and "revolution".
  - How often do we observe the so called paradigm shift? Connection with self-organized criticality: sandpile model, punctuated equilibrium?