Mining Graphs and Tensors

Christos Faloutsos
CMU

Thank you!

- Charlie Van Loan
- Lenore Mullin
- Frank Olken

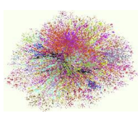
Outline

- Introduction – Motivation
- Problem#1: Patterns in static graphs
- Problem#2: Patterns in tensors / time evolving graphs
- Problem#3: Which tensor tools?
- Problem#4: Scalability
- Conclusions

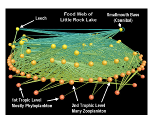
Motivation

➡ Data mining: ~ find patterns (rules, outliers)
- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
- Problem#3: What tools to use?
- Problem#4: Scalability to GB, TB, PB?

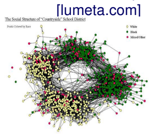
Graphs - why should we care?



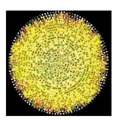
Internet Map
[lumeta.com]



Food Web
[Martinez '91]



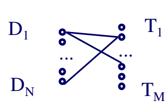
Friendship Network
[Moody '01]



Protein Interactions
[genomebiology.com]

Graphs - why should we care?

- IR: bi-partite graphs (doc-terms)



- web: hyper-text graph
- ... and more:

CMU SCS

Graphs - why should we care?

- network of companies & board-of-directors members
- 'viral' marketing
- web-log ('blog') news propagation
- computer network security: email/IP traffic and anomaly detection
- ....

CMU SCS

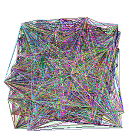
Outline

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
  - Degree distributions
  - Eigenvalues
  - Triangles
  - weights
- Problem#2: How do they evolve?
- …

CMU SCS

Problem #1 - network and graph mining



- How does the Internet look like?
- How does the web look like?
- What is 'normal'/'abnormal'?
- which patterns/laws hold?

CMU SCS

Graph mining

- Are real graphs random?

NSF tensors 2009 C. Faloutsos 10

CMU SCS

Laws and patterns

- Are real graphs random?
- A: NO!!
  - Diameter
  - in- and out- degree distributions
  - other (surprising) patterns

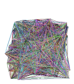
NSF tensors 2009 C. Faloutsos 11

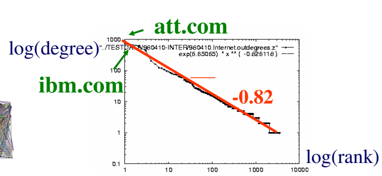
CMU SCS

Solution#1.1

- Power law in the degree distribution [SIGCOMM99]

internet domains



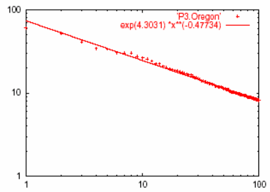


NSF tensors 2009 C. Faloutsos 12

CMU SCS

Solution#1.2: Eigen Exponent *E*

Eigenvalue



Rank of decreasing eigenvalue

Exponent = slope

E = -0.48

May 2001

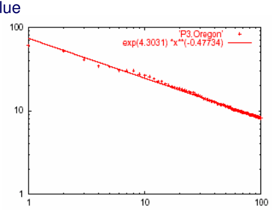
- A2: power law in the eigenvalues of the adjacency matrix

NSF tensors 2009 C. Faloutsos 13

CMU SCS

Solution#1.2: Eigen Exponent *E*

Eigenvalue



Rank of decreasing eigenvalue

Exponent = slope

E = -0.48

May 2001

- [Mihail, Papadimitriou ’02]: slope is ½ of rank exponent

NSF tensors 2009 C. Faloutsos 14

CMU SCS

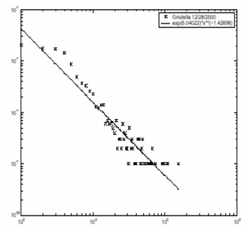
But:

How about graphs from other domains?

NSF tensors 2009 C. Faloutsos 15

CMU SCS

The Peer-to-Peer Topology



[Jovanovic+]

- Count versus degree
- Number of adjacent peers follows a power-law

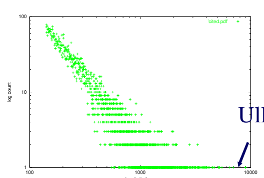
NSF tensors 2009 C. Faloutsos 16

CMU SCS

More settings w/ power laws:

citation counts: (*citeseer.nj.nec.com* 6/2001)

log(count)



log(#citations)

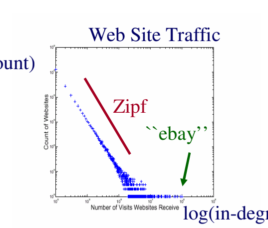
NSF tensors 2009 C. Faloutsos 17

CMU SCS

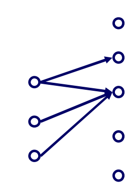
More power laws:

- web hit counts [w/ A. Montgomery]

log(count)



log(in-degree)

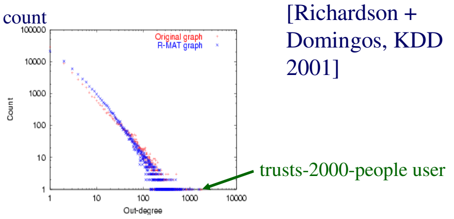


NSF tensors 2009 C. Faloutsos 18

CMU SCS

epinions.com

count



(out) degree

- who-trusts-whom [Richardson + Domingos, KDD 2001]

NSF tensors 2009 C. Faloutsos 19

CMU SCS

Outline

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
  - Degree distributions
  - Eigenvalues
  - Triangles
  - weights
- Problem#2: How do they evolve?
- …

NSF tensors 2009 C. Faloutsos 20

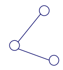
CMU SCS

How about triangles?

NSF tensors 2009 C. Faloutsos 21

CMU SCS

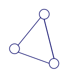
Solution# 1.3: Triangle 'Laws'



- Real social networks have a lot of triangles

CMU SCS

Triangle 'Laws'

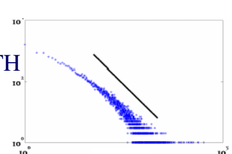


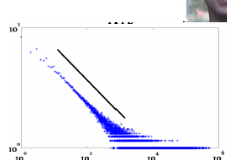
- Real social networks have a lot of triangles
  - Friends of friends are friends
- Any patterns?

CMU SCS

Triangle Law: #1
[Tsourakakis ICDM 2008]

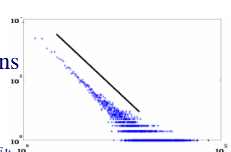
HEP-TH





ASN

Epinions

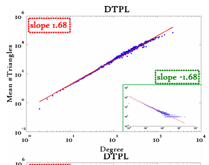


X-axis: # of Triangles
a node participates in
Y-axis: count of such nodes

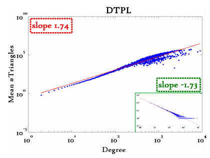
CMU SCS

Triangle Law: #2
[Tsourakakis ICDM 2008]

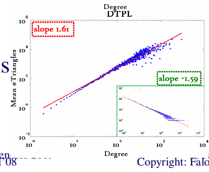
Reuters



SN



Epinions



X-axis: degree
Y-axis: mean # triangles
Notice: slope ~ degree exponent (insets)

CMU SCS

Triangle Law: Computations
[Tsourakakis ICDM 2008]

But: triangles are expensive to compute
(3-way join; several approx. algos)
Q: Can we do that quickly?

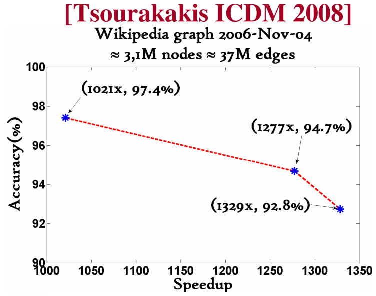
CMU SCS

Triangle Law: Computations
[Tsourakakis ICDM 2008]

But: triangles are expensive to compute
(3-way join; several approx. algos)
Q: Can we do that quickly?
A: Yes!
#triangles = 1/6 Sum (λ_i^3)
(and, because of skewness, we only need the top few eigenvalues!

Triangle Law: Computations

[Tsourakakis ICDM 2008]



1000x+ speed-up, high accuracy

Outline

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
  - Degree distributions
  - Eigenvalues
  - Triangles
  - weights
- Problem#2: How do they evolve?
- …

How about weighted graphs?

- A: even more 'laws'!

CMU SCS

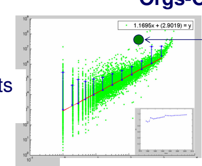
Solution#1.4: fortification

Q: How do the weights of nodes relate to degree?

CMU SCS

Solution#1.4: fortification: Snapshot Power Law

- At any time, total incoming weight of a node is proportional to in-degree with PL exponent 'iw':
  - *i.e. 1.01* < iw < 1.26, *super-linear*
- ***More donors, even more $***



CMU SCS

Motivation

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
  - Diameter
  - GCC, and NLCC
  - Blogs, linking times, cascades
- Problem#3: Tensor tools?
- …

CMU SCS

Problem#2: Time evolution

- with Jure Leskovec (CMU/MLD)

- and Jon Kleinberg (Cornell – sabb. @ CMU)

NSF tensors 2009 C. Faloutsos 34

CMU SCS

Evolution of the Diameter

- Prior work on Power Law graphs hints at **slowly growing diameter**:
  - diameter ~ O(log N)
  - diameter ~ O(log log N)
- What is happening in real data?

NSF tensors 2009 C. Faloutsos 35

CMU SCS

Evolution of the Diameter

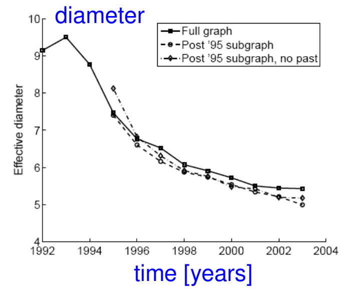
- Prior work on Power Law graphs hints at **slowly growing diameter**:
  - diameter ~ O(log N)
  - diameter ~ O(log log N)
- What is happening in real data?
- Diameter **shrinks** over time

NSF tensors 2009 C. Faloutsos 36

CMU SCS

Diameter – ArXiv citation graph

- Citations among physics papers
- 1992 –2003
- One graph per year

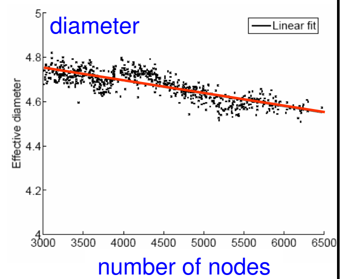


NSF tensors 2009 C. Faloutsos 37

CMU SCS

Diameter – “Autonomous Systems”

- Graph of Internet
- One graph per day
- 1997 – 2000

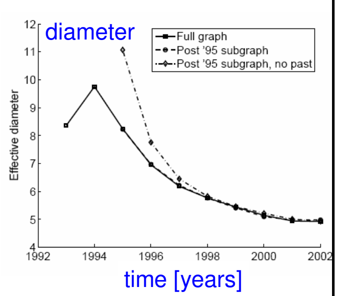


NSF tensors 2009 C. Faloutsos 38

CMU SCS

Diameter – “Affiliation Network”

- Graph of collaborations in physics – authors linked to papers
- 10 years of data

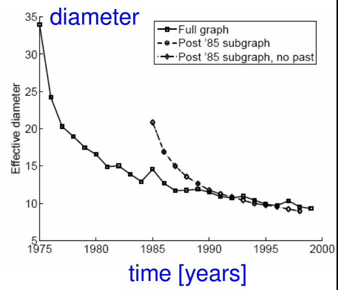


NSF tensors 2009 C. Faloutsos 39

CMU SCS

Diameter – "Patents"

- Patent citation network
- 25 years of data



NSF tensors 2009 C. Faloutsos 40

CMU SCS

Temporal Evolution of the Graphs

- N(t) … nodes at time t
- E(t) … edges at time t
- Suppose that

  N(t+1) = 2 * N(t)
- Q: what is your guess for

  E(t+1) =? 2 * E(t)

NSF tensors 2009 C. Faloutsos 41

CMU SCS

Temporal Evolution of the Graphs

- N(t) … nodes at time t
- E(t) … edges at time t
- Suppose that

  N(t+1) = 2 * N(t)
- Q: what is your guess for

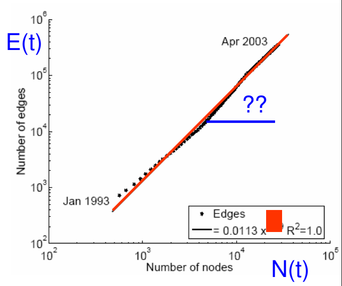
  E(t+1) =? 2 * E(t)
- A: over-doubled!
  - But obeying the ``Densification Power Law''

NSF tensors 2009 C. Faloutsos 42

CMU SCS

Densification – Physics Citations

- Citations among physics papers
- 2003:
  - 29,555 papers, 352,807 citations

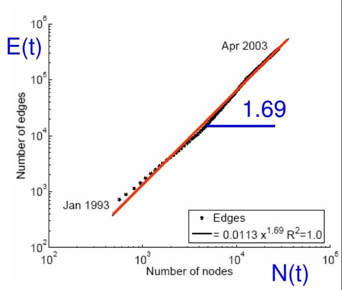


NSF tensors 2009 C. Faloutsos 43

CMU SCS

Densification – Physics Citations

- Citations among physics papers
- 2003:
  - 29,555 papers, 352,807 citations

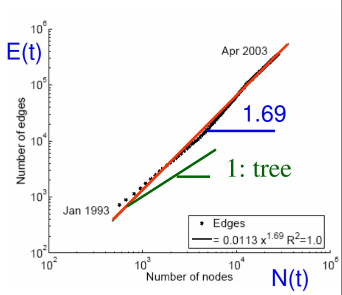


NSF tensors 2009 C. Faloutsos 44

CMU SCS

Densification – Physics Citations

- Citations among physics papers
- 2003:
  - 29,555 papers, 352,807 citations

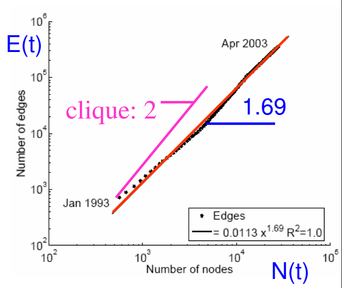


NSF tensors 2009 C. Faloutsos 45

CMU SCS

Densification – Physics Citations

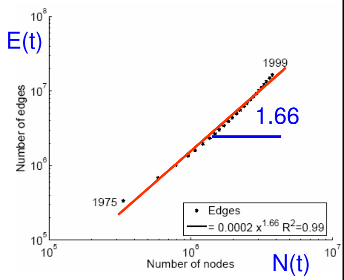
- Citations among physics papers
- 2003:
  - 29,555 papers, 352,807 citations



CMU SCS

Densification – Patent Citations

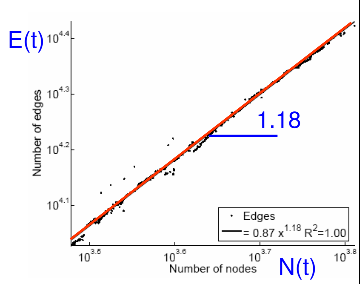
- Citations among patents granted
- 1999
  - 2.9 million nodes
  - 16.5 million edges
- Each year is a datapoint



CMU SCS

Densification – Autonomous Systems

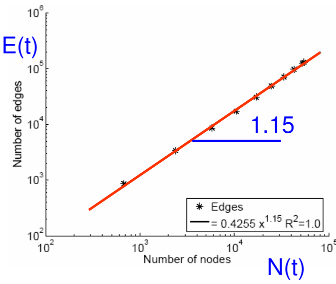
- Graph of Internet
- 2000
  - 6,000 nodes
  - 26,000 edges
- One graph per day



CMU SCS

Densification – Affiliation Network

- Authors linked to their publications
- 2002
  - 60,000 nodes
    - 20,000 authors
    - 38,000 papers
  - 133,000 edges



NSF tensors 2009 C. Faloutsos 49

CMU SCS

Motivation

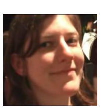
Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
  - Diameter
  - ➡ GCC, and NLCC
  - Blogs, linking times, cascades
- Problem#3: Tensor tools?
- …

NSF tensors 2009 C. Faloutsos 50

CMU SCS

More on Time-evolving graphs




M. McGlohon, L. Akoglu, and C. Faloutsos *Weighted Graphs and Disconnected Components: Patterns and a Generator. SIG-KDD* 2008

NSF tensors 2009 C. Faloutsos 51

CMU SCS

Observation 1: Gelling Point

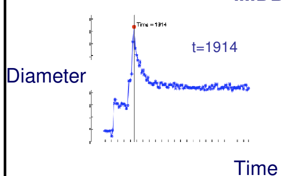
Q1: How does the GCC emerge?

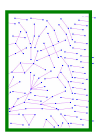
CMU SCS

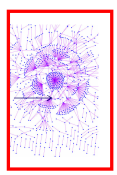
Observation 1: Gelling Point

- Most real graphs display a gelling point
- After gelling point, they exhibit typical behavior. This is marked by a spike in diameter.

IMDB







CMU SCS

Observation 2: NLCC behavior

Q2: How do NLCC's emerge and join with the GCC?

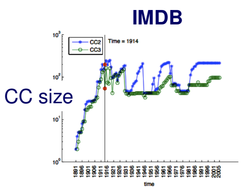
(``NLCC'' = non-largest conn. components)

– Do they continue to grow in size?
– or do they shrink?
– or stabilize?

CMU SCS

Observation 2: NLCC behavior

- After the gelling point, the GCC takes off, but NLCC's remain ~constant (actually, oscillate).



CMU SCS

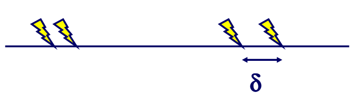
How do new edges appear?

[LBKT'08] [Microscopic Evolution of Social Networks](#)
Jure Leskovec, Lars Backstrom, Ravi Kumar, Andrew Tomkins.
(ACM KDD), 2008.

CMU SCS

How do edges appear in time? [LBKT'08]

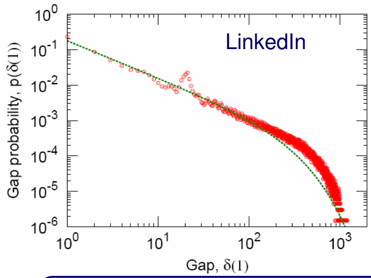
Edge gap δ(d): inter-arrival time between d^{th} and $d+1^{st}$ edge



What is the PDF of **δ?**
Poisson**?**

CMU SCS

How do edges appear in time? [LBKT'08]



Edge gap δ(d): inter-arrival time between d^{th} and $d+1^{st}$ edge

$$p_g(\delta(d);\alpha,\beta) \propto \delta(d)^{-\alpha} e^{-\beta\delta(d)}$$

1-58

CMU SCS

Motivation

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
  - Diameter
  - GCC, and NLCC
  - ➡ linking times, blogs, cascades
- Problem#3: Tensor tools?
- …

NSF tensors 2009 C. Faloutsos 59

CMU SCS

Blog analysis

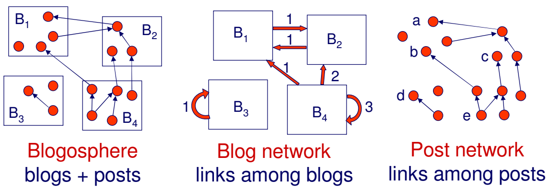
- with Mary McGlohon (CMU)
- Jure Leskovec (CMU)
- Natalie Glance (now at Google)
- Mat Hurst (now at MSR)

[SDM'07]

NSF tensors 2009 C. Faloutsos 60

CMU SCS

Cascades on the Blogosphere

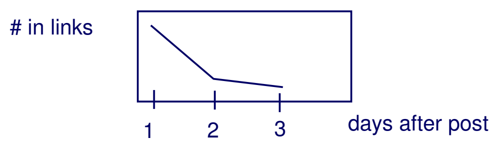


Q1: popularity-decay of a post?
Q2: degree distributions?

NSF tensors 2009 C. Faloutsos 61

CMU SCS

Q1: popularity over time



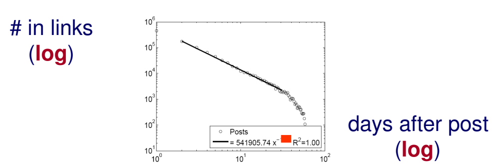
Post popularity drops-off – exponentially?

NSF tensors 2009 C. Faloutsos

Days after post

CMU SCS

Q1: popularity over time



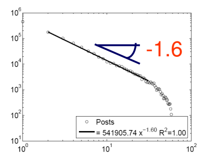
Post popularity drops-off – exponentially?
POWER LAW!
Exponent?

NSF tensors 2009 C. Faloutsos

Days after post

CMU SCS

Q1: popularity over time



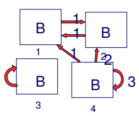
Post popularity drops-off – exponentially?
POWER LAW!
Exponent? -1.6 (close to -1.5: Barabasi's stack model)

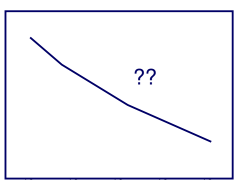
NSF tensors 2009 C. Faloutsos Days after post

CMU SCS

Q2: degree distribution

44,356 nodes, 122,153 edges. Half of blogs belong to largest connected component.



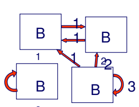


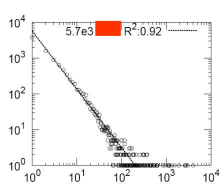
NSF tensors 2009 C. Faloutsos 65

CMU SCS

Q2: degree distribution

44,356 nodes, 122,153 edges. Half of blogs belong to largest connected component.



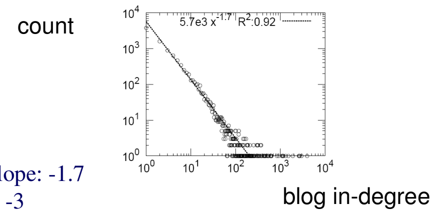


NSF tensors 2009 C. Faloutsos 66

CMU SCS

Q2: degree distribution

44,356 nodes, 122,153 edges. Half of blogs belong to largest connected component.



in-degree slope: -1.7
out-degree: -3
'rich get richer'

NSF tensors 2009 C. Faloutsos 67

CMU SCS

Motivation

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
- ➡ Problem#3: What tools to use?
  - PARAFAC/Tucker, CUR, MDL
  - generation: Kronecker graphs
- Problem#4: Scalability to GB, TB, PB?

NSF tensors 2009 C. Faloutsos 68

CMU SCS

Tensors for time evolving graphs

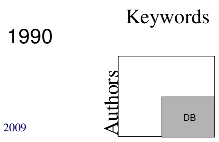
- [Jimeng Sun+ KDD'06]
- [" , SDM'07]
- [CF, Kolda, Sun, SDM'07 tutorial]

NSF tensors 2009 C. Faloutsos 69

CMU SCS

Social network analysis

- Static: find community structures

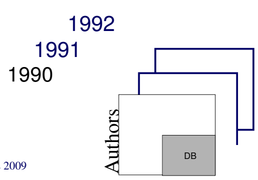


NSF tensors 2009 70

CMU SCS

Social network analysis

- Static: find community structures

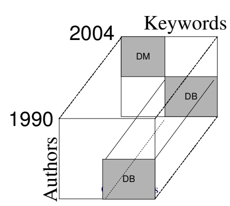


NSF tensors 2009 71

CMU SCS

Social network analysis

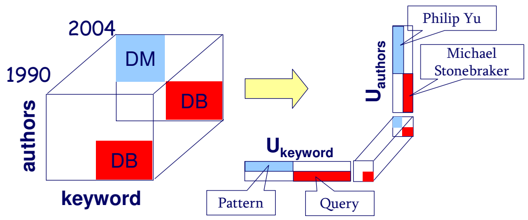
- Static: find community structures
- Dynamic: monitor community structure evolution; spot abnormal individuals; abnormal time-stamps



NSF tensors 2009 72

CMU SCS

Application 1: Multiway latent semantic indexing (LSI)



- Projection matrices specify the clusters
- Core tensors give cluster activation level

CMU SCS

Bibliographic data (DBLP)

- Papers from VLDB and KDD conferences
- Construct 2nd order tensors with yearly windows with <author, keywords>
- Each tensor: 4584×3741
- 11 timestamps (years)

CMU SCS

Multiway LSI

Authors	Keywords	Year
michael carey, michael stonebraker, h. jagadish, hector garcia-molina	queri,parallel,optimization,concurr, objectorient	1995
surajit chaudhuri,mitch cherniack,michael stonebraker,ugur etintemel	distribut,systems,view,storage,servic,process,cache	2004
jiawei han,jian pei,philip s. yu, jianyong wang,charu c. aggarw[illegible]	streams,pattern,support, cluster, [illegible] gener,queri	2004

DB

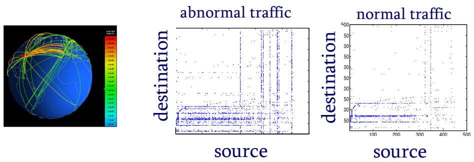
DM

- Two groups are correctly identified: Databases and Data mining
- People and concepts are drifting over time

CMU SCS

Network forensics

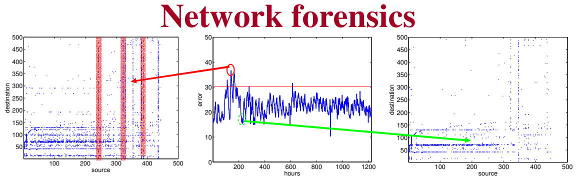
- Directional network flows
- A large ISP with 100 POPs, each POP 10Gbps link capacity [Hotnets2004]
  - 450 GB/hour with compression
- Task: Identify abnormal traffic pattern and find out the cause



NSF t (with Prof. Hui Zhang, Dr. Jimeng Sun, Dr. Yinglian Xie) 76

CMU SCS

Network forensics



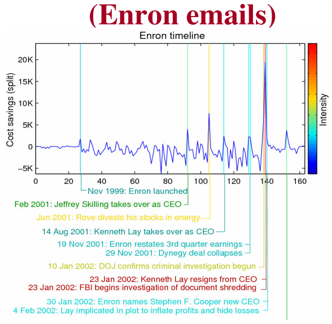
Abnormal traffic | Reconstruction error over time | Normal traffic

- Reconstruction error gives indication of anomalies.
- Prominent difference between normal and abnormal ones is mainly due to the unusual scanning activity (confirmed by the campus admin).

NSF tensors 2009 C. Faloutsos 77

CMU SCS

MDL mining on time-evolving graph (Enron emails)



NSF tensors 2009

GraphScope [w. Jimeng Sun, Spiros Papadimitriou and Philip Yu, KDD'07]

Motivation

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
- Problem#3: What tools to use?
  - PARAFAC/Tucker, CUR, MDL
  - generation: Kronecker graphs
- Problem#4: Scalability to GB, TB, PB?

Problem#3: Tools - Generation

- Given a growing graph with count of nodes N_1, N_2, …
- Generate a realistic sequence of graphs that will obey all the patterns

Problem Definition

- Given a growing graph with count of nodes N_1, N_2, …
- Generate a realistic sequence of graphs that will obey all the patterns
  - Static Patterns
    - Power Law Degree Distribution
    - Power Law eigenvalue and eigenvector distribution
    - Small Diameter
  - Dynamic Patterns
    - Growth Power Law
    - Shrinking/Stabilizing Diameters

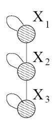
CMU SCS

Problem Definition

- Given a growing graph with count of nodes N_1, N_2, …
- Generate a realistic sequence of graphs that will obey all the patterns
- **Idea: Self-similarity**
  - Leads to power laws
  - Communities within communities
  - …

CMU SCS

Kronecker Product – a Graph



1	1	0
1	1	1
0	1	1

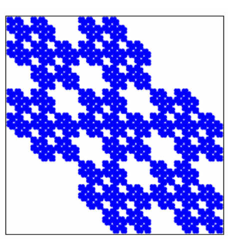
G_1

Adjacency matrix

CMU SCS

Kronecker Product – a Graph

- Continuing multiplying with G_1 we obtain G_4 and so on …

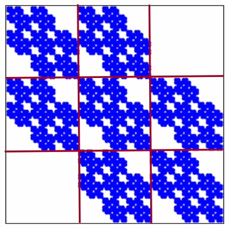


G_4 adjacency matrix

CMU SCS

Kronecker Product – a Graph

- Continuing multiplying with G_1 we obtain G_4 and so on …

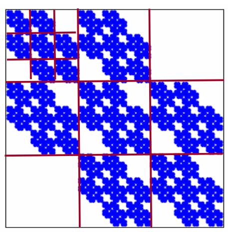


G_4 adjacency matrix

CMU SCS

Kronecker Product – a Graph

- Continuing multiplying with G_1 we obtain G_4 and so on …



G_4 adjacency matrix

CMU SCS

Properties:

- We can PROVE that
  - Degree distribution is multinomial ~ power law
  - Diameter: constant
  - Eigenvalue distribution: multinomial
  - First eigenvector: multinomial
- See [Leskovec+, PKDD'05] for proofs

CMU SCS

Problem Definition

- Given a growing graph with nodes N_1, N_2, …
- Generate a realistic sequence of graphs that will obey all the patterns
  - Static Patterns
    - ✔ Power Law Degree Distribution
    - ✔ Power Law eigenvalue and eigenvector distribution
    - ✔ Small Diameter
  - Dynamic Patterns
    - ✔ Growth Power Law
    - ✔ Shrinking/Stabilizing Diameters
- First and only generator for which we can **prove** all these properties

NSF tensors 2009 C. Faloutsos 88

CMU SCS

skip

Stochastic Kronecker Graphs

- Create $N_1 \times N_1$ probability matrix P_1
- Compute the k^{th} Kronecker power P_k
- For each entry p_{uv} of P_k include an edge (u,v) with probability p_{uv}

0.4	0.2
0.1	0.3

P_1

Kronecker multiplication →

0.16	0.08	0.08	0.04
0.04	0.12	0.02	0.06
0.04	0.02	0.12	0.06
0.01	0.03	0.03	0.09

P_k

→ flip biased coins

Instance Matrix G_2

NSF tensors 2009 C. Faloutsos 89

CMU SCS

Experiments

- How well can we match real graphs?
  - Arxiv: physics citations:
    - 30,000 papers, 350,000 citations
    - 10 years of data
  - U.S. Patent citation network
    - 4 million patents, 16 million citations
    - 37 years of data
  - Autonomous systems – graph of internet
    - Single snapshot from January 2002
    - 6,400 nodes, 26,000 edges
- We show both static and temporal patterns

NSF tensors 2009 C. Faloutsos 90

CMU SCS

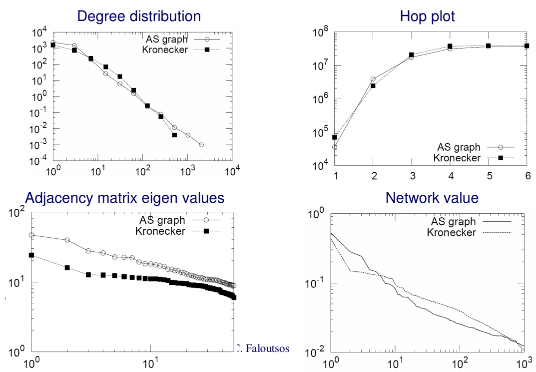
(Q: how to fit the parm's?)

A:

- Stochastic version of Kronecker graphs +
- Max likelihood +
- Metropolis sampling
- [Leskovec+, ICML'07]

CMU SCS

Experiments on real AS graph



CMU SCS

Conclusions

- Kronecker graphs have:
  - All the static properties
    - ✓Heavy tailed degree distributions
    - ✓Small diameter
    - ✓Multinomial eigenvalues and eigenvectors
  - All the temporal properties
    - ✓Densification Power Law
    - ✓Shrinking/Stabilizing Diameters
  - We can formally prove these results

CMU SCS

How to generate realistic tensors?

- A: 'RTM' [Akoglu+, ICDM'08]
  - do a tensor-tensor Kronecker product
  - resulting tensors (= time evolving graphs) have bursty addition of edges, over time.

NSF tensors 2009 C. Faloutsos 94

CMU SCS

Motivation

Data mining: ~ find patterns (rules, outliers)

- Problem#1: How do real graphs look like?
- Problem#2: How do they evolve?
- Problem#3: What tools to use?
- Problem#4: Scalability to GB, TB, PB?

NSF tensors 2009 C. Faloutsos 95

CMU SCS

Scalability

- How about if graph/tensor does not fit in core?
- How about handling huge graphs?

NSF tensors 2009 C. Faloutsos 96

CMU SCS

Scalability

- How about if graph/tensor does not fit in core?
- ['MET': Kolda, Sun, ICMD'08, best paper award]
- How about handling huge graphs?

NSF tensors 2009 C. Faloutsos 97

CMU SCS

Scalability

- Google: > 450,000 processors in clusters of ~2000 processors each [Barroso, Dean, Hölzle, *"Web Search for a Planet: The Google Cluster Architecture"* IEEE Micro 2003]
- Yahoo: 5Pb of data [Fayyad, KDD'07]
- Problem: machine failures, on a daily basis
- How to parallelize data mining tasks, then?
- A: map/reduce – hadoop (open-source clone) http://hadoop.apache.org/

NSF tensors 2009 C. Faloutsos 98

CMU SCS

2' intro to hadoop

- master-slave architecture; n-way replication (default n=3)
- 'group by' of SQL (in parallel, fault-tolerant way)
- e.g, find histogram of word frequency
  - compute local histograms
  - then merge into global histogram

```
select course-id, count(*)
from ENROLLMENT
group by course-id
```

NSF tensors 2009 C. Faloutsos 99

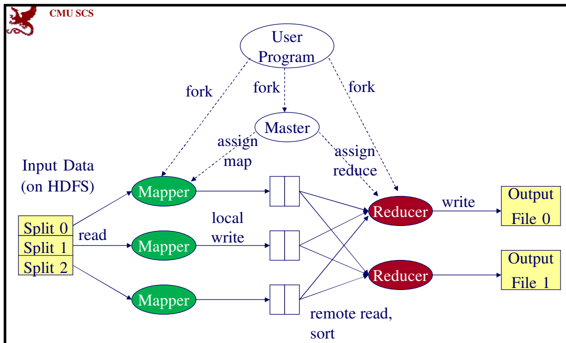
CMU SCS

2’ intro to hadoop

- master-slave architecture; n-way replication (default n=3)
- ‘group by’ of SQL (in parallel, fault-tolerant way)
- e.g, find histogram of word frequency
  - compute local histograms
  - then merge into global histogram

select course-id, count(*)   reduce
from ENROLLMENT
group by course-id   map

CMU SCS



By default: 3-way replication;
Late/dead machines: ignored, **transparently** (!)

CMU SCS

D.I.S.C.

- ‘Data Intensive Scientific Computing’ [R. Bryant, CMU]
  - ‘big data’
  - http://www.cs.cmu.edu/~bryant/pubdir/cmu-cs-07-128.pdf

CMU SCS

E.g.: self-* and DCO systems @ CMU

- >200 nodes
- target: 1 PetaByte
- Greg Ganger +:
  - www.pdl.cmu.edu/SelfStar
  - www.pdl.cmu.edu/DCO

CMU SCS

OVERALL CONCLUSIONS

- Graphs/tensors pose a wealth of fascinating problems
- self-similarity and power laws work, when textbook methods fail!
- New patterns (densification, fortification, -1.5 slope in blog popularity over time
- New generator: Kronecker
- Scalability / cloud computing -> PetaBytes

CMU SCS

References

- Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan *Fast Random Walk with Restart and Its Applications* ICDM 2006, Hong Kong.
- Hanghang Tong, Christos Faloutsos *Center-Piece Subgraphs: Problem Definition and Fast Solutions,* KDD 2006, Philadelphia, PA
- T. G. Kolda and J. Sun. *Scalable Tensor Decompositions for Multi-aspect Data Mining*. In: ICDM 2008, pp. 363-372, December 2008.

CMU SCS

References

- Jure Leskovec, Jon Kleinberg and Christos Faloutsos *Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations* KDD 2005, Chicago, IL. ("Best Research Paper" award).
- Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos *Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication* (ECML/PKDD 2005), Porto, Portugal, 2005.

NSF tensors 2009 C. Faloutsos 106

CMU SCS

References

- Jure Leskovec and Christos Faloutsos, *Scalable Modeling of Real Graphs using Kronecker Multiplication*, ICML 2007, Corvallis, OR, USA
- Shashank Pandit, Duen Horng (Polo) Chau, Samuel Wang and Christos Faloutsos *NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks* WWW 2007, Banff, Alberta, Canada, May 8-12, 2007.
- Jimeng Sun, Dacheng Tao, Christos Faloutsos *Beyond Streams and Graphs: Dynamic Tensor Analysis,* KDD 2006, Philadelphia, PA

NSF tensors 2009 C. Faloutsos 107

CMU SCS

References

- Jimeng Sun, Yinglian Xie, Hui Zhang, Christos Faloutsos. *Less is More: Compact Matrix Decomposition for Large Sparse Graphs*, SDM, Minneapolis, Minnesota, Apr 2007. [pdf]
- Jimeng Sun, Spiros Papadimitriou, Philip S. Yu, and Christos Faloutsos, *GraphScope: Parameter-free Mining of Large Time-evolving Graphs* ACM SIGKDD Conference, San Jose, CA, August 2007

NSF tensors 2009 C. Faloutsos 108

CMU SCS

THANK YOU!

Contact info:

www. cs.cmu.edu /~christos

(w/ papers, datasets, code, etc)

Funding sources:

- NSF IIS-0705359, IIS-0534205, DBI-0640543
- LLNL, PITA, IBM, INTEL

NSF tensors 2009 C. Faloutsos 109