Shift of research interest in Computer Sciences over last 50 years: What citation analysis reflects?

Animesh Mukherjee Department of Computer Science and Engg. Indian Institute of Technology, Kharagpur

**Collaborators: Tanmoy Chakraborty, Sandipan Sikdar, Niloy Ganguly** 



# **CITATION NETWORKS**



# MOTIVATION

 Are you planning to pursue a research career in computer science?

• Confused about which filed to opt for?

 We propose to search for answers to some of these questions

# DATASET

• Citation graph constructed from DBLP dump

• Has bibliographic information till 2009

• Historically used for the Arnetminer project (Tang et al. 2008)

• Reference: http://arnetminer.org/citation, named as DBLP-Citation-network V4

# **RAW FACTS**





### **NUMBER OF PUBLICATIONS**



No of new publications per year is directly proportional to the existing publications

## **AVG. INWARD CITATIONS**



### **INWARD AND OUTWARD CITATIONS**



New inward citations is proportional to existing inward citations New outward citations is proportional to existing outward citations

### FIELDS OF COMPUTER SCIENCE

No.	Subject	Abbreviation	% of papers
1	Artificial Intelligence	AT	15.30
2	Algorithms and Theory	ALCO	14.00
2.	Networking	NW	8.63
4	Detabage	DP	8 10
4. F	Databases	DET	0.12
э.	Parallel Computing	DIST	7.03
6.	Hardware & Architecture	ARC	7.29
7.	Software Engineering	SE	6.40
8.	Machine Learning and Pattern Recognition	ML	6.09
9.	Scientific Computing	SC	4.02
10.	Bioinformatics	BIO	3.88
	& Computational Biology		
11.	Human-Computer	HCI	3.42
	Interaction		
12.	Multimedia	MUL	3.34
13.	Graphics	GRP	3.32
14.	Computer Vision	CV	3.03
15.	Data Mining	DM	3.02
16.	Programming Languages	PL	3.00
17.	Security and Privacy	SEC	2.94
18.	Information Retrieval	IR	2.26
19.	Natural Language	NLP	2.11
	and Speech		
20.	World Wide Web	WWW	1.76
21.	Computer Education	EDU	1.67
22.	Operating Systems	OS	1.07
23.	Real Time	RT	0.90
	Embedded Systems		
24.	Simulation	SIM	0.14

## **AUTHORITY OF A FIELD**

• For field-field networks (in our case, 24 fields) we compute *Inwardness* as follows

$$In(f_i^t) = \sum_{j \neq i} w_{j \to i}^t$$

where,

 $w_{j \to i}^{t} = \frac{C_{j \to i}^{t}}{P_{i}^{t}}$   $C_{j \to i}^{t} = \# \text{ of citations received by the papers of field } i$ from field j  $P_{i}^{t} = \# \text{ of papers in field } i$ 

 $1 \le t \le 3$  (current year + next 3 years)

### **OVERALL AUTHORITY OF FIELDS**

Abbreviation	% of papers	Average Inwardness				
		60-69	70-79	80-89	90-99	00-08
AI	15.30	0.02	0.67	4.94	5.14	3.29
ALGO	14.09	4.13	4.49	3.39	2.12	0.55
NW	8.63	0.19	0.53	1.06	3.42	1.76
DB	8.12	3.75	3.67	1.80	1.14	0.17
DIST	7.63	0.02	2.02	2.86	1.55	0.56
	- <b>2</b> 0					
ARC	7.29	0.41	2.49	2.29	1.12	1.04
SE	6.40	1.98	3.21	1.89	1.67	0.52
ML	6.09	0	0.43	2.51	2.97	2.62
SC	4.02	0	1.14	2.38	2.91	0.19
BIO	3.88	0	0	0.71	1.27	0.56
II.OI	0.40	0	0.00	1.05	0.05	1.00
HCI	3.42	0	0.03	1.65	2.05	1.39
MIIL	2.24	0	0.52	9.51	9.99	1.99
CPP	2 20	0	0.55	2.51	2.22	1.05
CV	3.02	0	0.00	1.00	2.03	1.07
DM	2.00	0	0.00	1.29	1.09	1.27
DM	3.02	0.41	0.27	1.00	1.00	1.02
FL	3.00	0.41	2.49	2.00	2.40	1.29
JD JD	2.94	0	0.00	3.00	2.00	1.59
	2.20	0	0.42	1.32	2.62	1.79
NLP	2.11	0	0.13	1.16	2.82	1.92
WWW	1.76	0	0	1.86	2.10	1.83
EDU	1.67	0	0	0.80	0.83	0.39
OS	1.07	0.31	1.73	1.39	1.98	1.20
RT	0.90	0	0.67	1.56	2.52	0.54
SIM	0.14	0	0.30	1.20	2.70	0.87

# **TIME-TRANSITION DIAGRAM**



- Rise in inwardness & decline near transition throughout
- Second ranked field emerges as the leader in the next window.
- Top fields undergoing a decline can **again lead later**  $\rightarrow$  (contradicting Bornholdt et al., *PRL*, 2011)

IEEE/ACM ASONAM 2013. [Acceptance rate: 13%], best paper nominee

# **REASONS: COLLABORATIONS**

### • For the top ranked fields, measure:

- Collaborative papers (papers with multiple authors)
- Multi-continental papers
- Diversity of a papers (average number of fields in which authors have worked)



# **REASON: HIGH IMPACT PAPERS**



Frac. of top and second rank fields among the 10% high impact papers

- 82% cases → fraction of top ranked field's papers declines and second ranked field rises at the transition point.
- Initial years → the sum of these two was nearly 1; current years
   → it has decreased to a point below 0.5 → Strong competition!!



Backup fields: Those fields that provide citations to other fields
In 75% cases, citation patterns from the backup fields decline at the transition period → citations get distributed among the fields.

IEEE/ACM ASONAM 2013. [Acceptance rate: 13%]

## **REASONS: SEMINAL PAPERS**

**Influence:** Influence
$$(p_i^t) = \sum_{p_j \in P^t} \frac{1}{d_{p_j}}$$

 $P^t$  = set of all papers that cite  $p_i$  within 3 years window and belong to the same field as  $p_i$ ;  $d_{p_i}$  = out-degree of  $p_j$ ;  $p_i$  and  $p_j$  does not have any author in common



- Rank the fields from among the10 % high influential papers
- 65% cases, most influential field emerges as the leader in the subsequent phase.

## **PUBLICATIONS IN THE CONTINENTS**



# **TEAM EFFORT**



Both team size and multi-continent papers increase over time

## **IMPACT OF THE CONTINENTS**



North America dominates others from the very beginning



- Frequent transitions
- Most of the cases, third fields emerge at top
- Behavioural patterns of the continents have a correspondence with the global behaviour that either lags or leads in time

# WHO INFLUENCES WHOM?

- Similarity( $\zeta$ ) = s / n; s = similarity pair, n = no. of year = 46
- lead(X, Y, t) = X took place t years before Y



- EU has much correlation with Global at t = 0
- NA has much correlation with Global at t = 1

# **FUNDING STATISTICS OF NSF**

		NSF		
Yrs	Inwardness	Proposal	Proposal	
	results	submitted	awarded	
03	AI/IR/NW	NW/AI/HCI	NW/ALGO/SE	
04	AI/IR/NW	AI/HCI/RT	RT/ARC/DIST	
05	AI/IR/NW	AI/ML/HCI	GRP/SE/ALGO	
06	IR/ML/AI	ML/ALGO/SEC	ALGO/SEC/ML	
07	ML/AI/ALGO	ALGO/ML/HCL	ALGO/HCI/SEC	
08	ML/AI/ALGO	ML/ALGO/SE	ALGO/ML/SE	

Top three fields based on (i) our prediction, (ii) proposal submission statistics and (iii) award statistics from 2003 to 2008

IEEE/ACM ASONAM 2013. [Acceptance rate: 13%], best paper nominee

# **FUND CORRELATIONS**

• Correlation( $\zeta$ ) = s / n; s = similarity pair, n = no of year = 46

- *lead(X, Y, t)* = X took place *t* years before Y
- lag(X, Y, t) = X took place t years after Y



# **INSIGHTS**

- Second ranked field always emerges as the lead in the next window. Top fields that undergo a decline can again lead later.
- North America, having largest overall impact, seems to regulate the overall research direction.
- Predicted results **perfectly correlates** with the proposal **submission** statistics, and **partially correlates** with **funds awarded**.
- **Future Objective:** Building **recommendation system** that can predict **forthcoming scientific research focus**!!

### HOW TO SEEK THE RIGHT TOOLKIT?





#### Interdisciplinary toolkit

## IN THE LINES OF GREAT THINKERS

"We are not students of some subject matter, but students of problems. And problems may cut right across the borders of any subject matter or discipline."

-- Karl Popper

"Interdisciplinary research is the only way to do research in current times."

– Fritjof Capra

## **CITATIONS/REFERENCES FROM OTHER FIELDS**



### **REFERENCE DIVERSITY INDEX**

RDI of a paper  $X_i = RDI(X_i) = -\sum_i p_i \log p_i$ 

where,  $p_i$  is the proportion of references of  $X_i$  citing the papers of field  $f_i$ 



### **RANKING FIELDS BASED ON RDI**



## **DRIFT OF CITATION DIVERSITY INDEX**

> CDI of a paper X<sub>i</sub> at time t<sub>i</sub> =  $CDI_{t_i}(X_i) = -\sum_j p_j \log p_j$ 

where,  $p_j$  is the proportion of citations received by  $X_i$  from the papers of field  $f_j$ 

Drift of CDI between two successive time windows =

$$\Delta_{t_{i}}(f_{i}) = CDI_{t_{i+1}}(f_{i}) - CDI_{t_{i}}(f_{i})$$

# **SPIKES IN CDI**



### **MEMBERSHIP DIVERSITY INDEX**

$$\blacktriangleright MDI \text{ of a field } f_i = MDI(f_i) = -\sum_j p_j \log p_j$$

where,  $p_j$  is the fraction of overlapped papers of field  $f_i$  belonging to the communities tagged as  $f_j$ 

Overlapping community detection: *SLPA* Based on intra-community citation density

(Xie et al., ICDM, 2011)

### **RANKING FIELDS BASED ON MDI**



### **EXTERNAL EVIDENCE: ATTRACTION INDEX**

$$\chi_f = \frac{n_{i+4} - n_i}{c_i}$$

 $n_i$ : number of unique authors from the beginning up to the year  $t_i$  publishing papers in field f

 $n_{i+4}$ : number of unique authors from the beginning up to the year  $t_{i+4}$  publishing papers in field f

 $c_i$ : number of publications in the time window  $(t_{i+4} - t_i)$ 

### **RANK BASED ON ATTRACTION INDEX**



### **CLASSIFICATION BASED ON ABOVE RANKS**



IEEE/ASE Socialcom 2013. [Acceptance rate: 9.9%]

### **EVOLUTIONARY LANDSCAPE**



### **IMPACT OF INTERDISCIPLINARITY**

### **CITATION MEASURES**

Statistics	Core	Interdisciplinary
	fields (CR)	fields(INT)
Avg. number of papers per field	21,846	28,091
Avg. number of citations received	1.343	2.433
by a paper		
Avg. number of citations received	1.349	1.8125
by an author		
Avg. number of citations received	7276	8382
by a venue (in terms of the papers		
published in that venue)		
Fraction of papers among the top	0.338	0.662
10% high impact papers		
Fraction of authors among the top	0.257	0.743
10% high impact authors		
Fraction of cross-citations	0.0800	0.06747
	$(CR \rightarrow INT)$	$(INT \rightarrow CR)$

### **IMPACT OF INTERDISCIPLINARITY**

### **TOP-TIER CONFERENCES**



# **CORE-PERIPHERY ANALYSIS**



# http://cnerg.org



