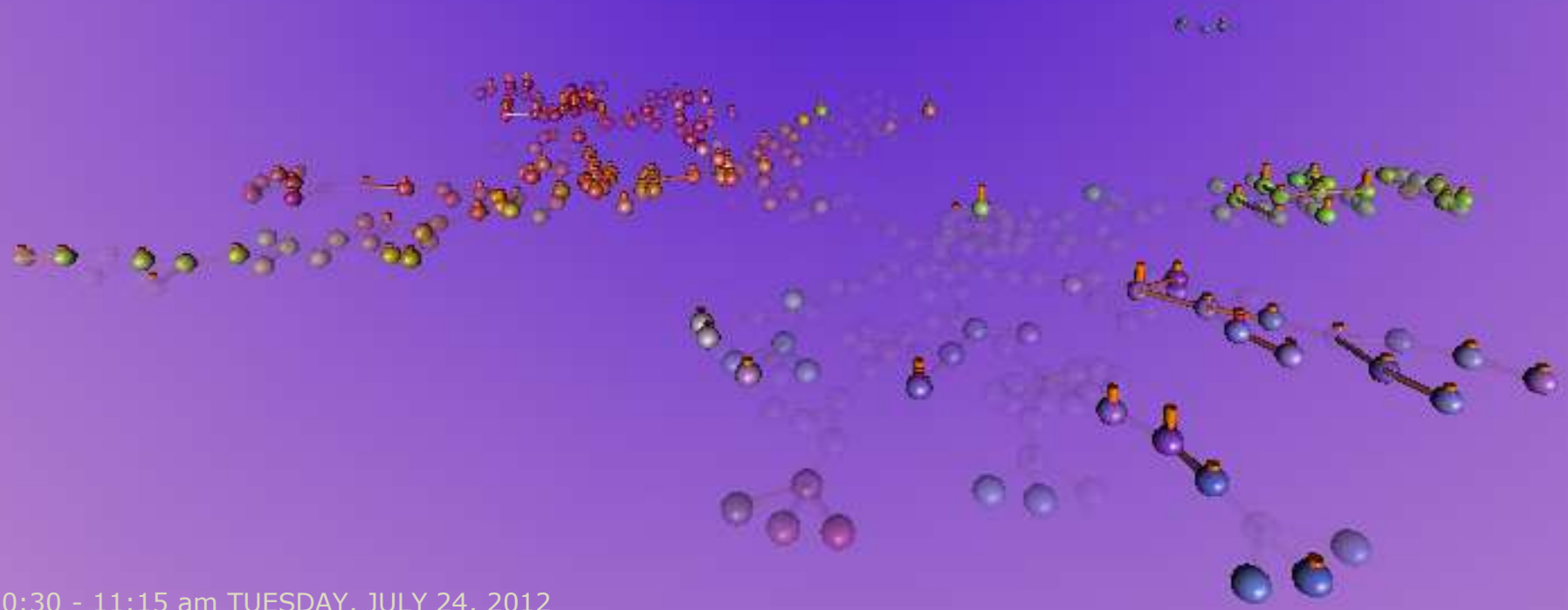
Detecting Potentially Transformative Research:
The Theory and Examples of Structural Variation

Chaomei Chen

Drexel University



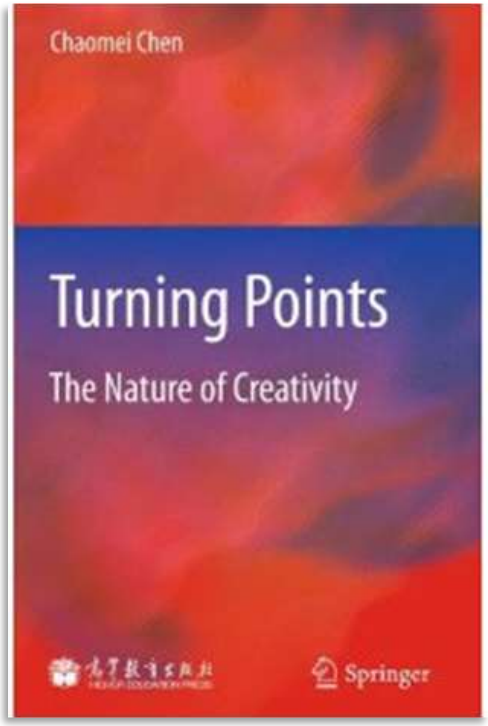
10:30 - 11:15 am TUESDAY, JULY 24, 2012
PORTFOLIO ANALYSIS SYMPOSIUM
NIH CAMPUS, BETHESDA, MD

Further Details

Email: chaomei.chen@drexel.edu

Chen, C. (2011) *Turning Points: The Nature of Creativity*. Springer. ISBN-10: 3642191592; ISBN-13: 978-3642191596

Chen, C. (2012) Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, **63**(3), 431-449.



The Problem

- How can we gain foresights of the development of scientific fields and disciplines?
- How can we get better prepared for the future based on the past and present?

Measuring the Impact in terms of Structural Variation

- How to measure the potential value of a stream of new ideas, or information in general, in terms of how it may ***alter*** the current knowledge of a scientific domain, which may range from specialties to disciplines
- How to identify factors that influence the ***diffusion*** of specific knowledge.

Extrinsic Indicators

- Citations received
- Citations increased fast
- Times viewed online
- Times downloaded
- Produced by reputable researchers
- Published in high-impact journals

……

However,

- The underlying reasons are not clear.
- One may have to wait for data to accumulate, e.g. getting citations.
- It is hard to act on such information along.
- Past may not be a good indicator of the future (at least in this context)

Intrinsic Indicators

- Novelty of a concept
- Novelty of a new association
- Strength of an association
- Centrality of a paper

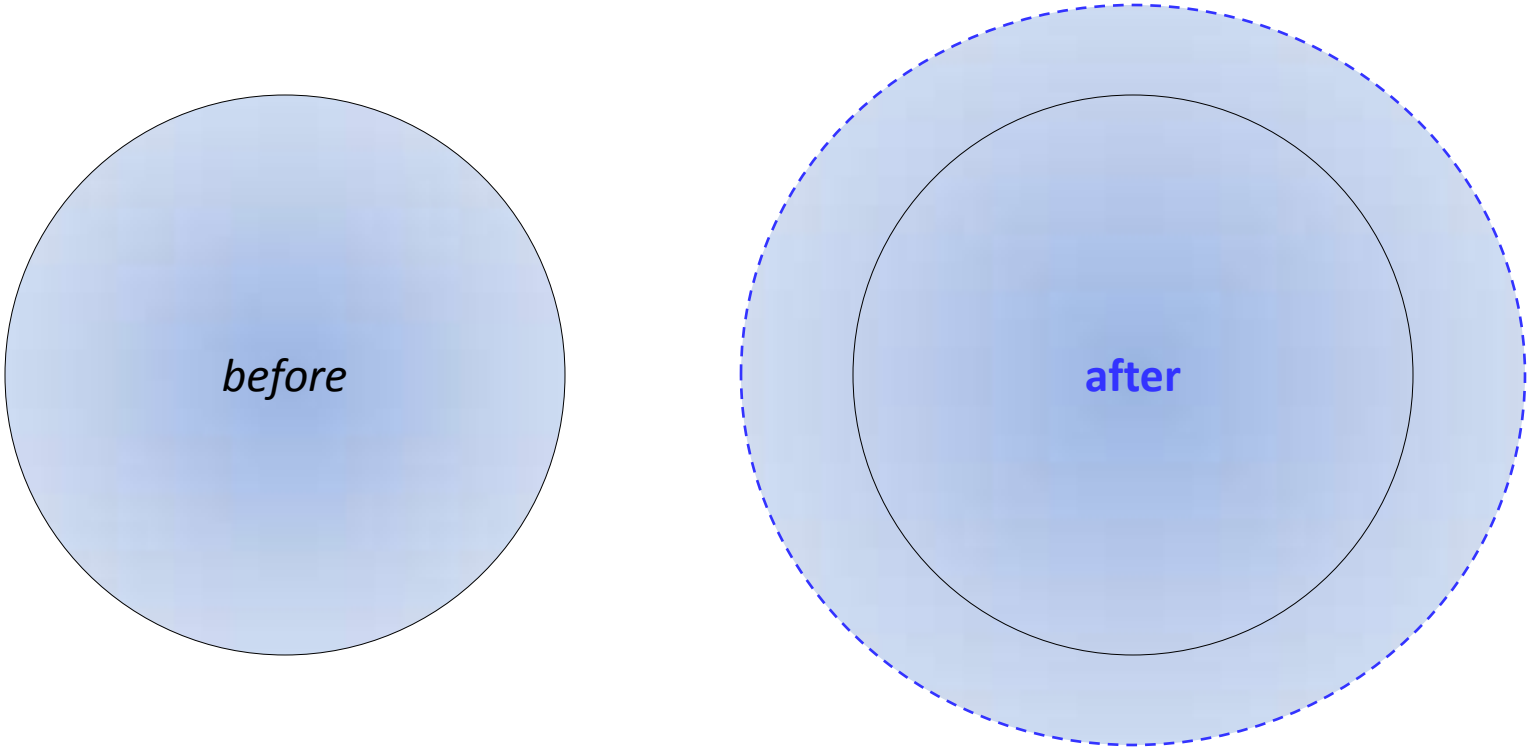
……

Advantages:

- Measure the substance of new ideas directly
- Reduce the necessary exposure time before making estimates
- Provide more information about the reasons for an impact or an anticipated impact
- Provide specific targets to act on

How Scientific Knowledge Grows

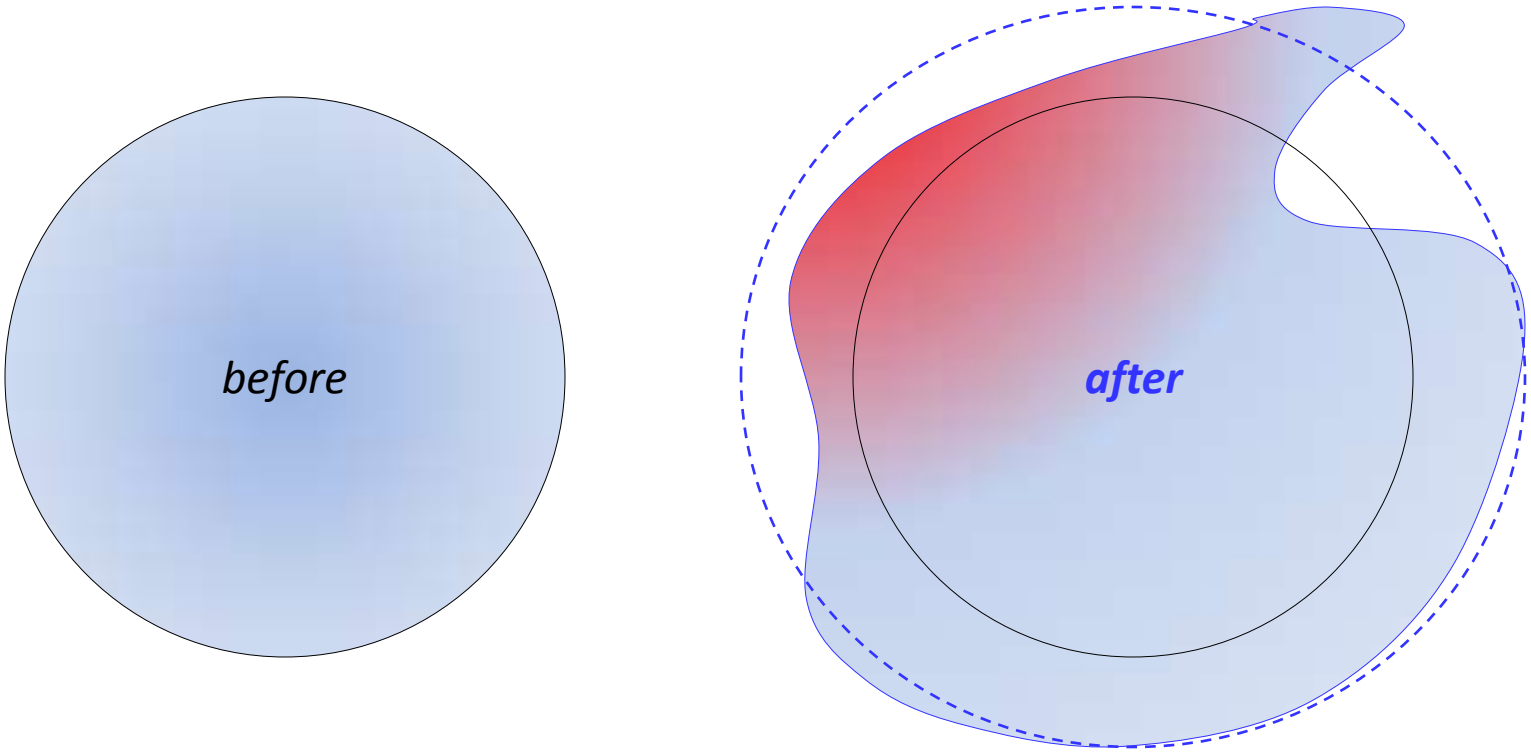
The Simplest Scenario

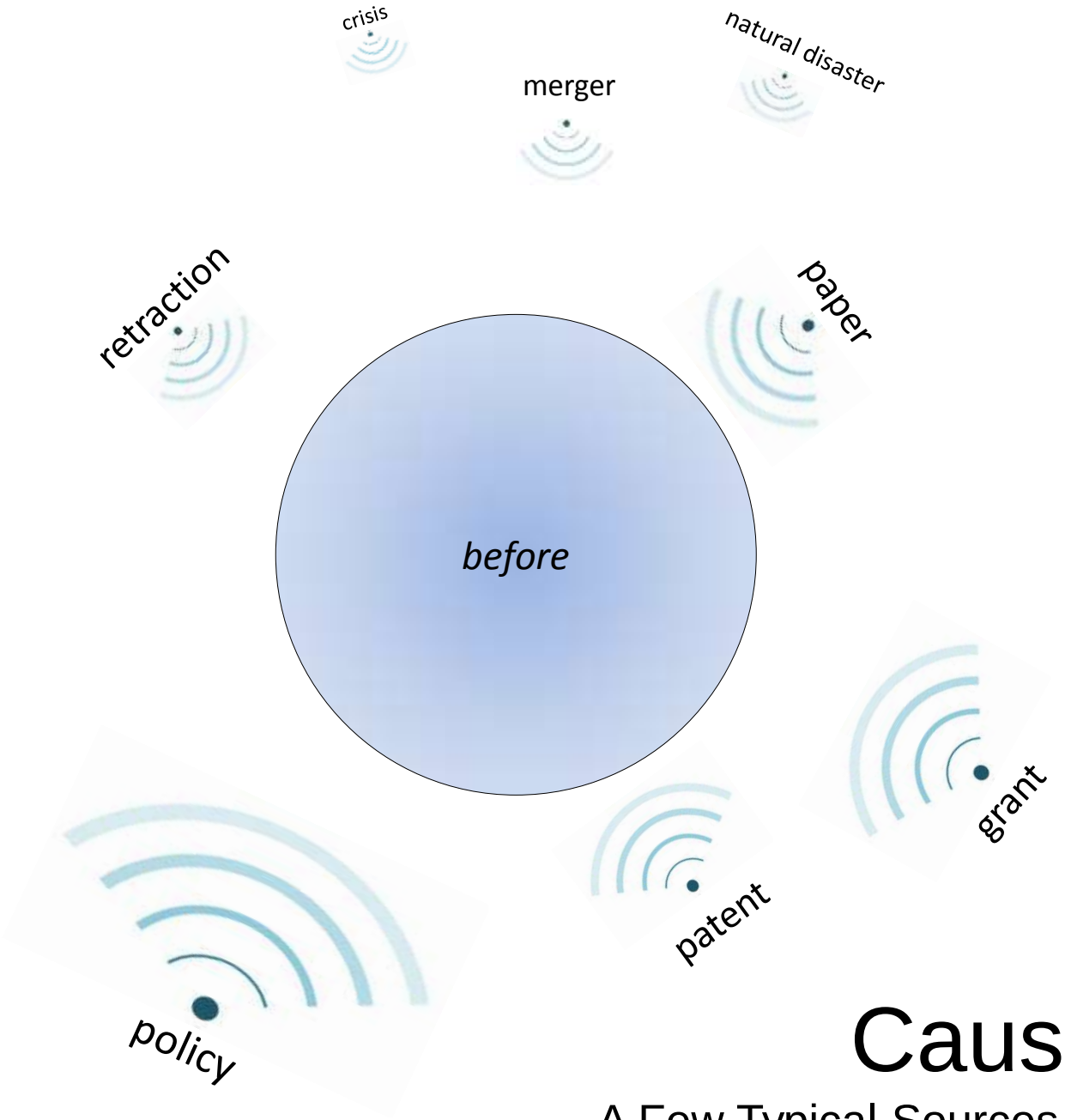


Transformative Growth:

Existing beliefs are altered and overturned

The name of the game in science



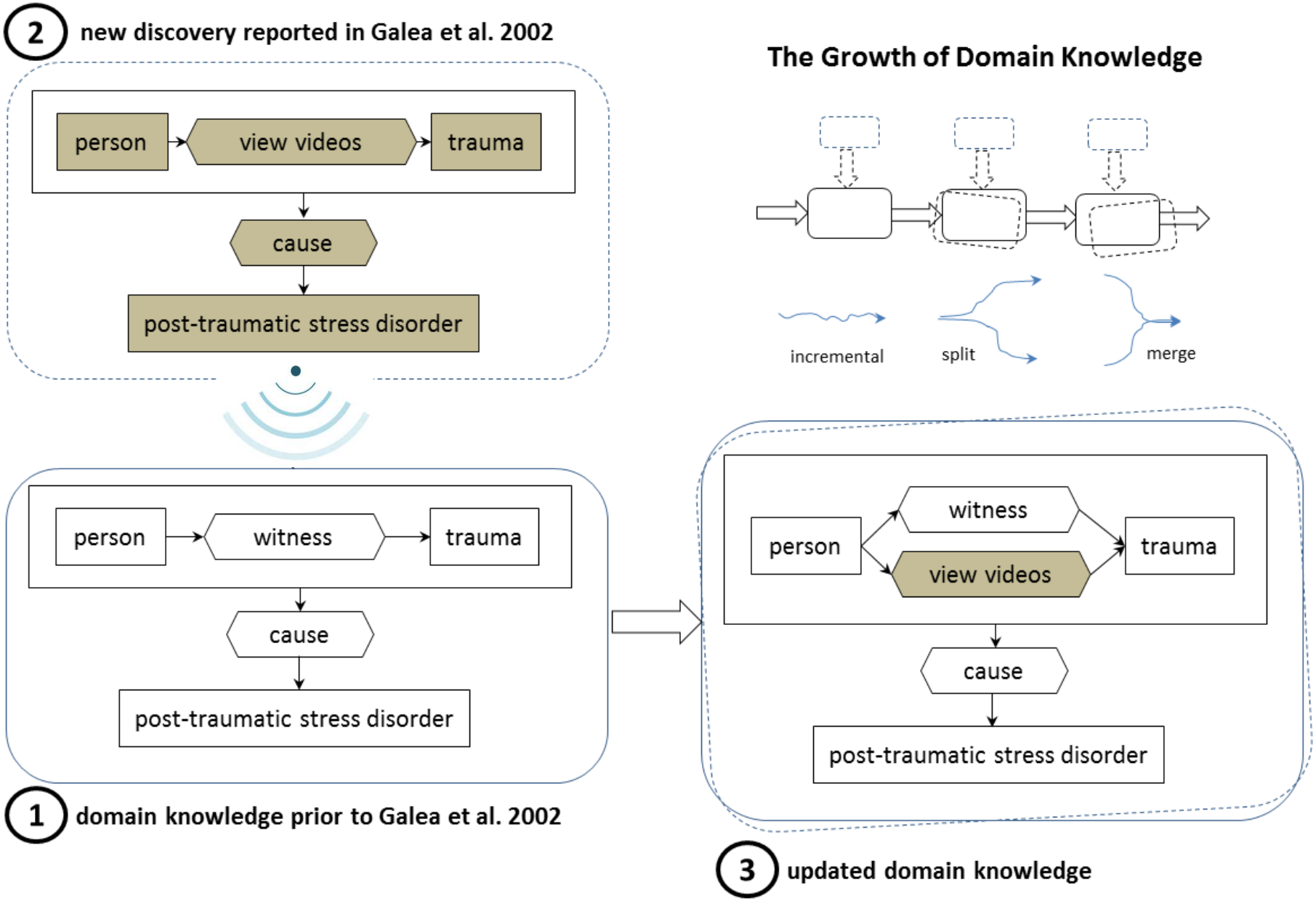


Cause and Effect

A Few Typical Sources of System-Level Impact

How Our Beliefs Are Changed by New Information

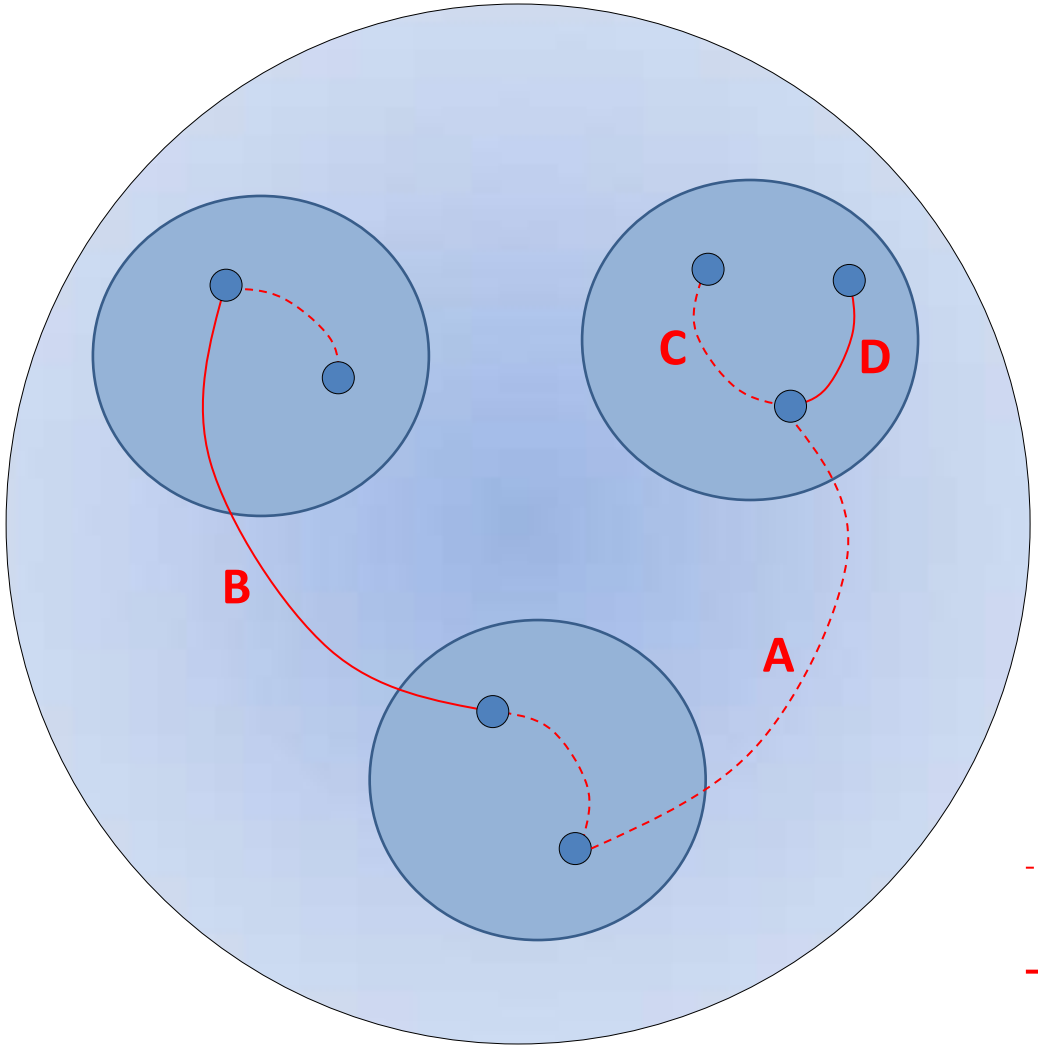
Example: Post-Traumatic Stress Disorder (PTSD)

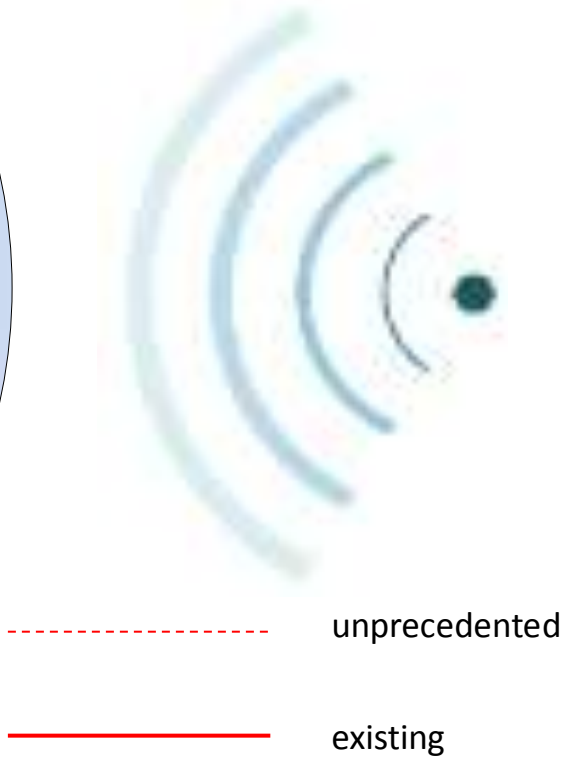


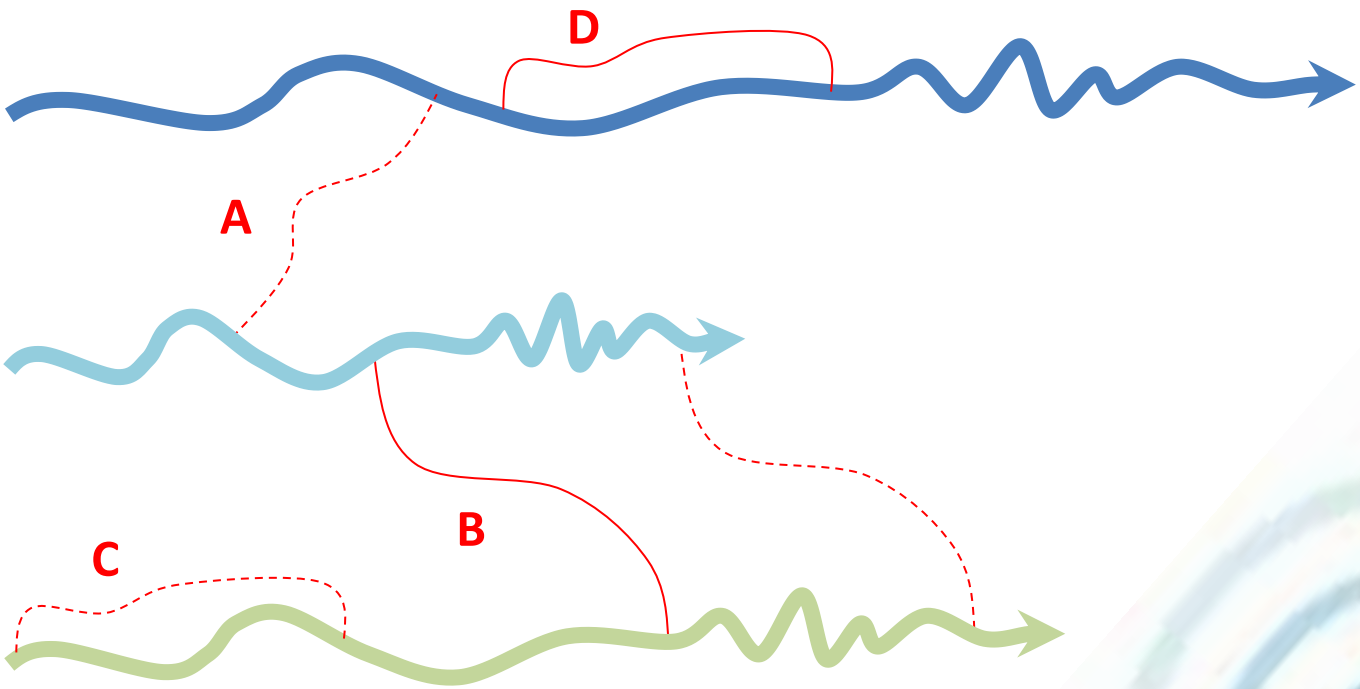
Measuring Transformative Potential in terms of Structural Variations

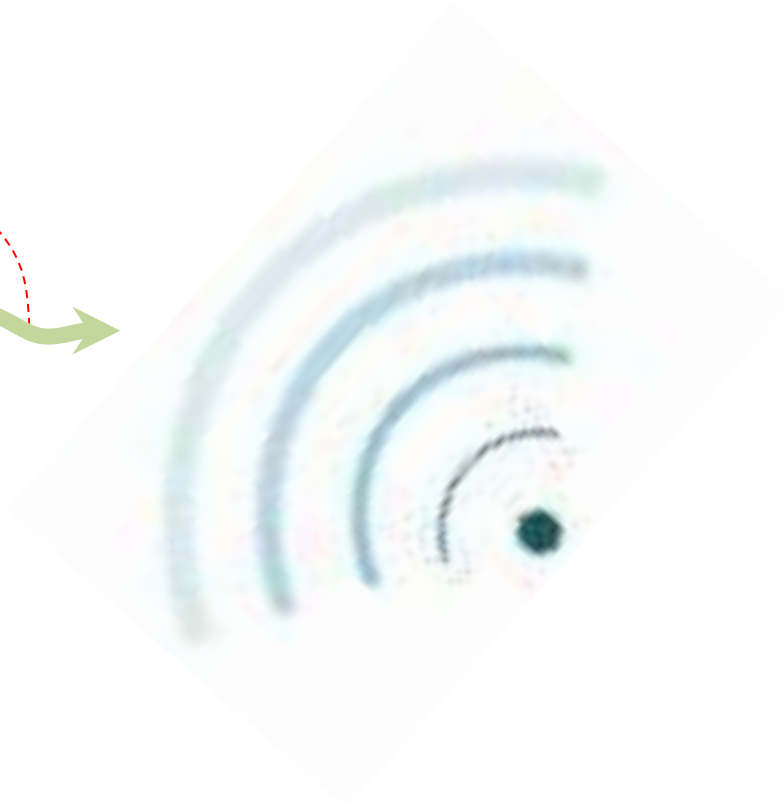
A) new between clusters; B) existed between clusters
C) new within cluster; D) existed within cluster

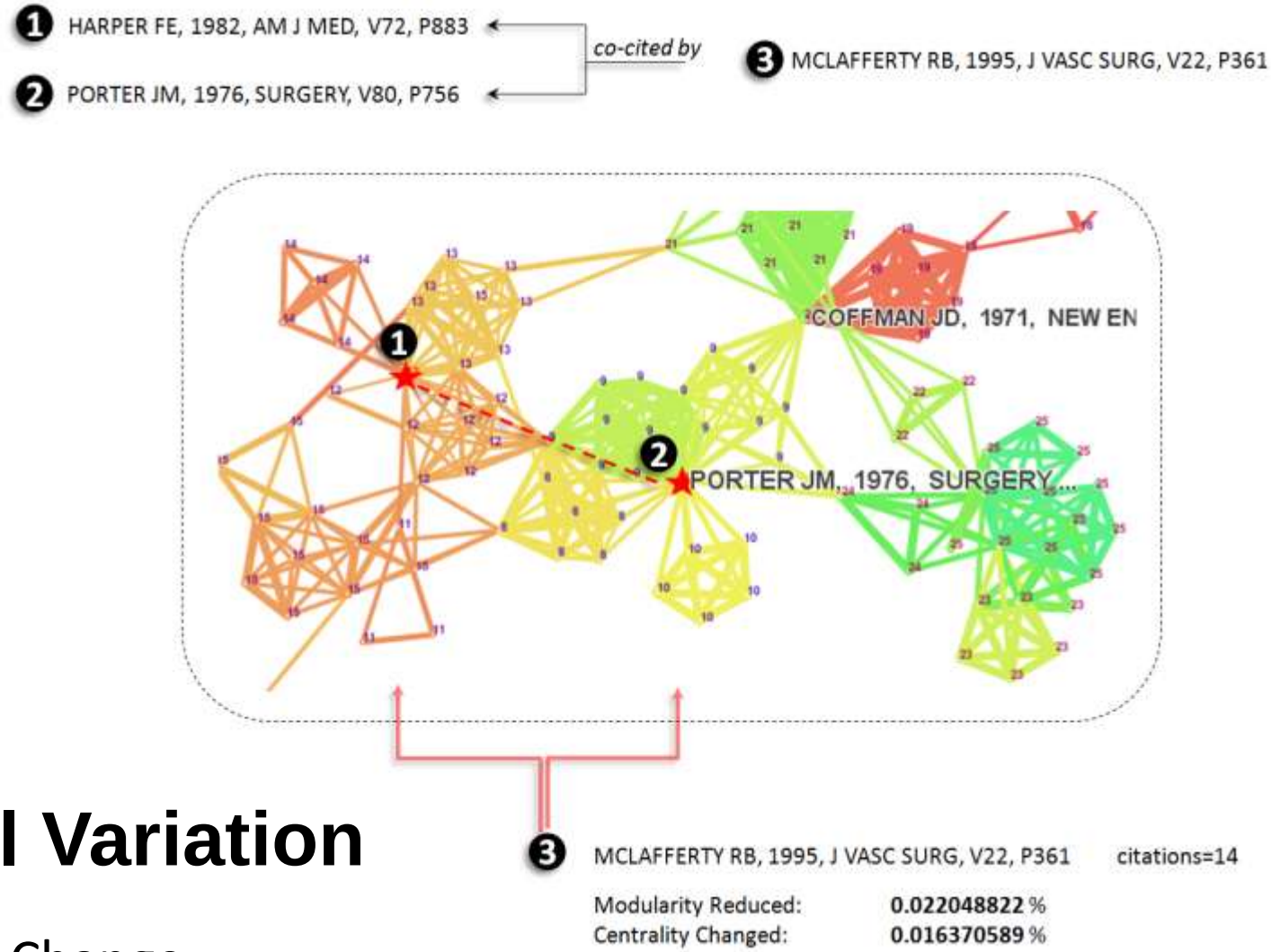
A, B =>> Cluster Linkage
A, B, C, D =>> Modularity; Centrality Divergence






D
A
B
C





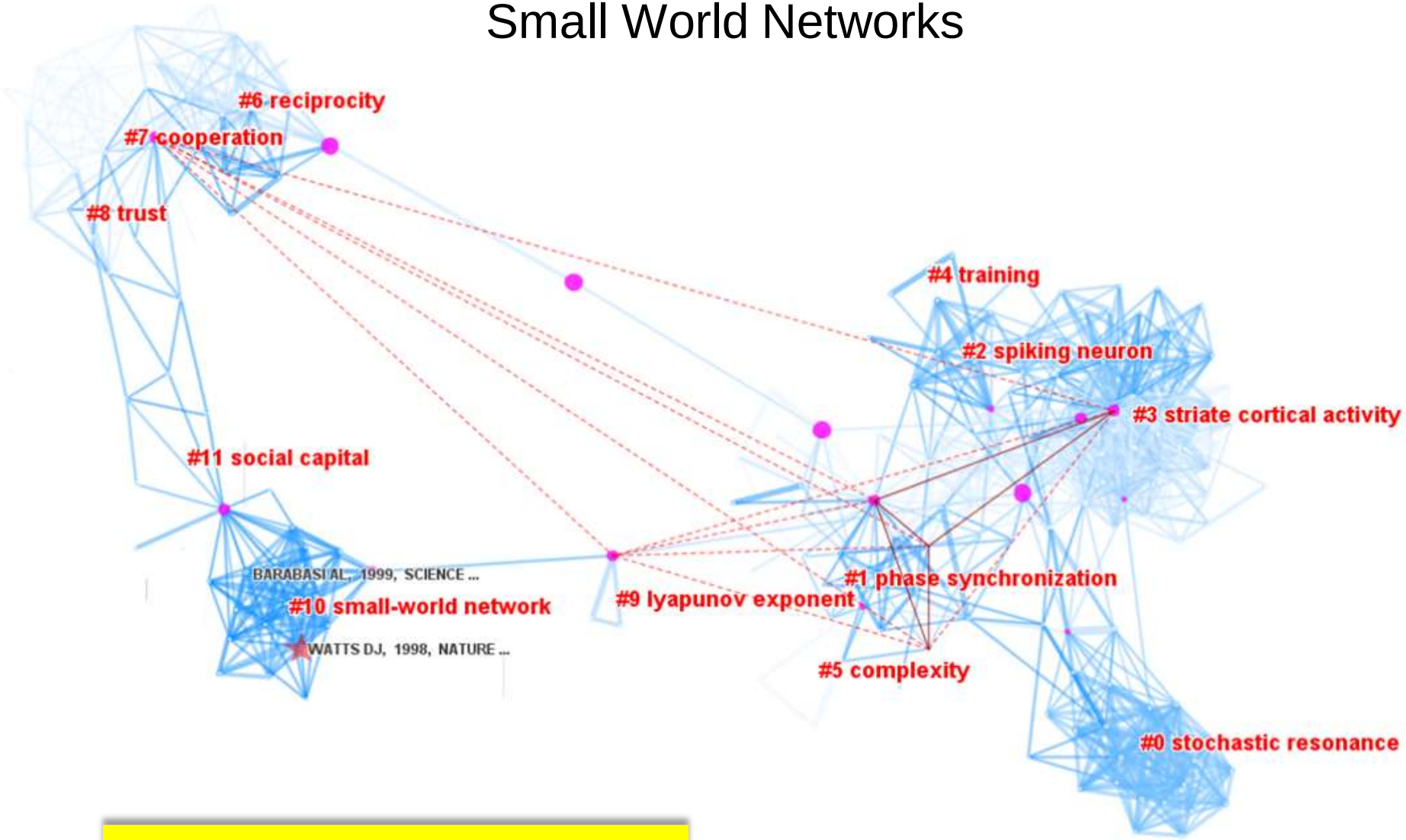
Structural Variation

- Modularity Change
- New Bridges between Clusters
- Centrality Divergence

Source: Chen, C. (2010) Measuring Structural Change in Networks due to New Information. *NATO IST-093/RWS-015 Workshop on Visualizing Networks: Coping with Change and Uncertainty*. Griffiss Institutes, Rome, NY. October 19-21, 2010.

Examples: Scientific Literature

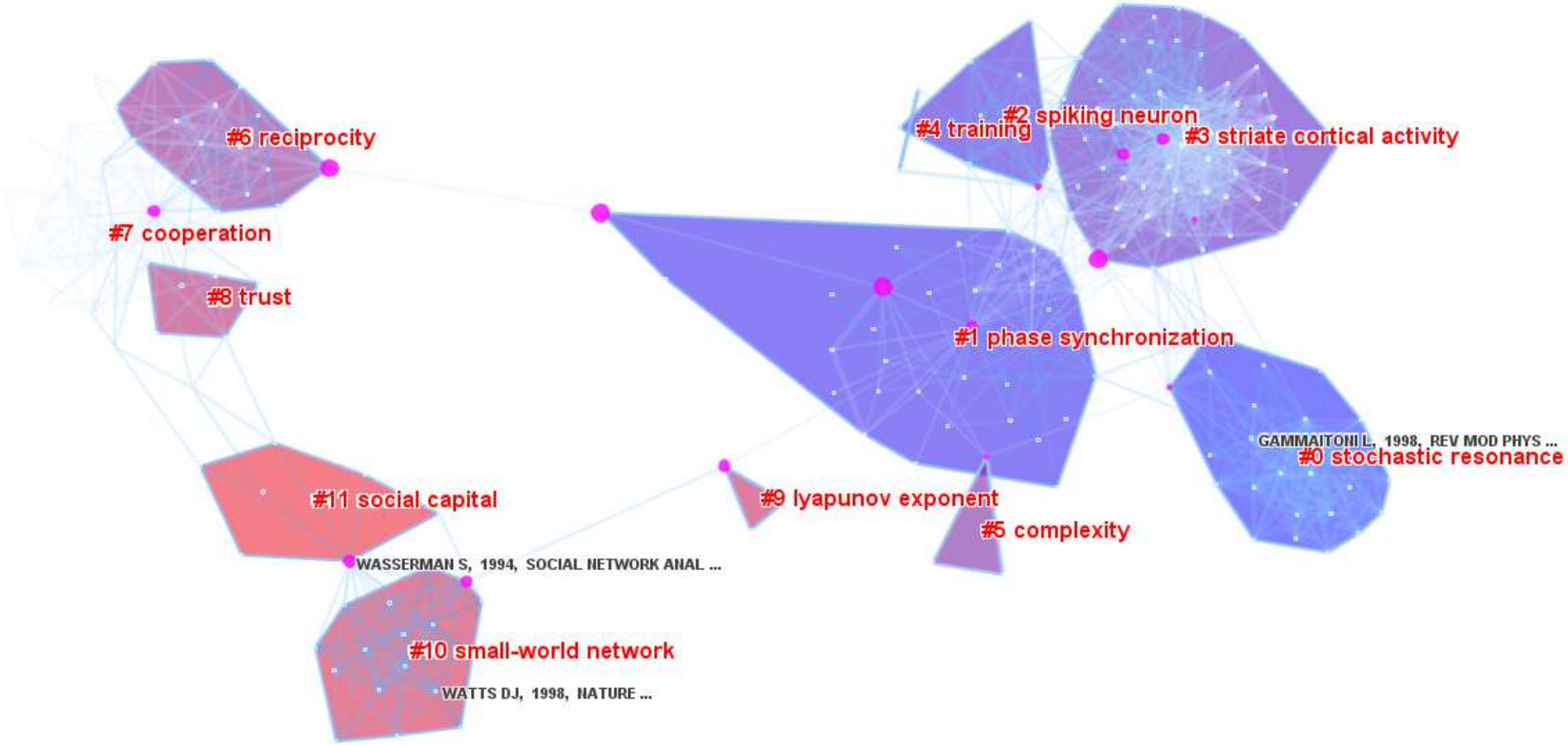
Small World Networks



New bridges made by Watts in 1998

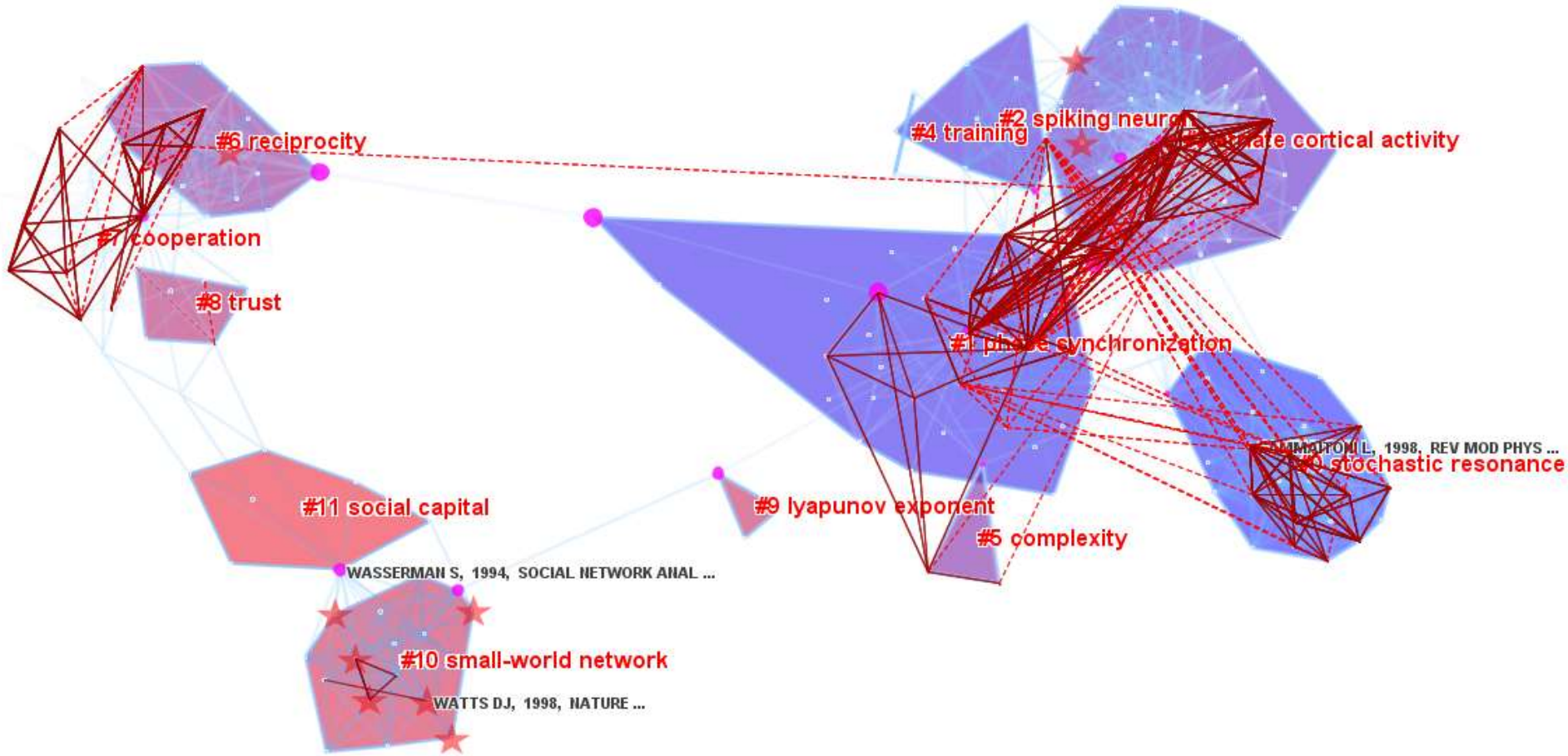
Examples: Scientific Literature

Small World Networks – Before Watts 1998



Examples: Scientific Literature

Small World Networks – After Watts 1998



A network of co-cited references derived from 5,135 articles published on small-world networks between 1990-2010. The network of 205 references and 1,164 co-citation links is divided into 12 clusters with a modularity of 0.6537 and the mean silhouette of 0.811. The red lines are made by the top-15 articles measured by the centrality variation rate.

Predictive Effects of Structural Variation Variables on Future Citations

The Strengths of Boundary-Spanning Signals versus Extrinsic Signals

Goals: 1) Follow the guidance of intrinsic signals to transformative research

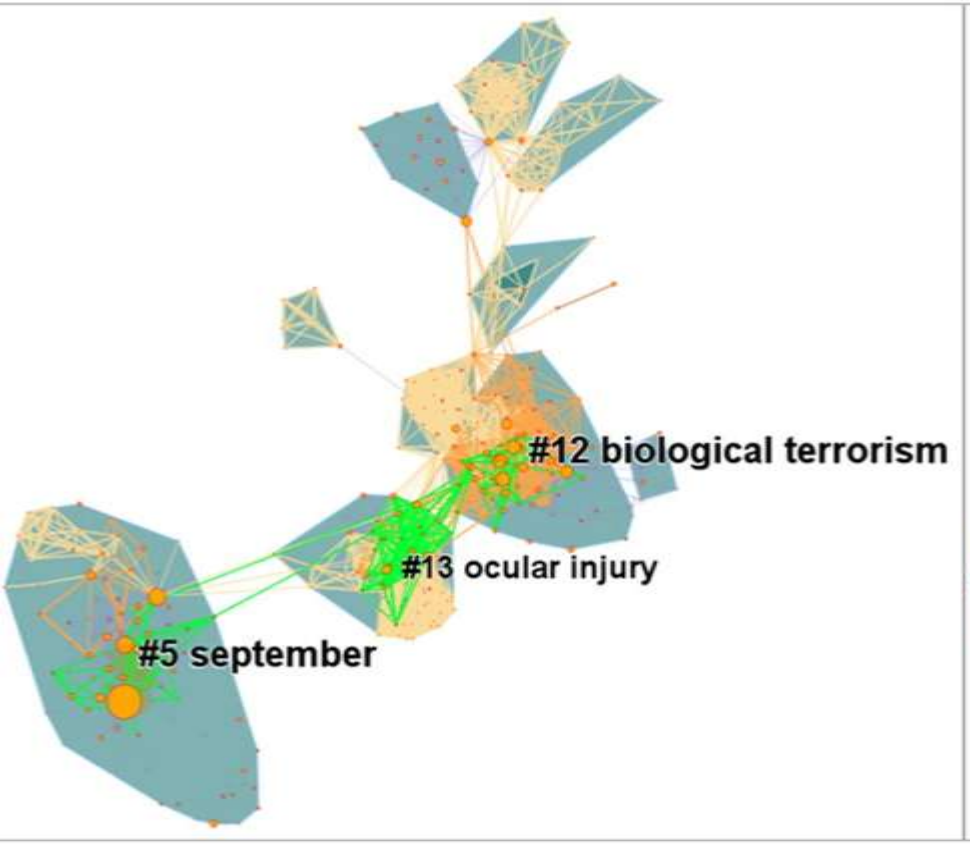
2) Anticipate where new breakthroughs may emerge

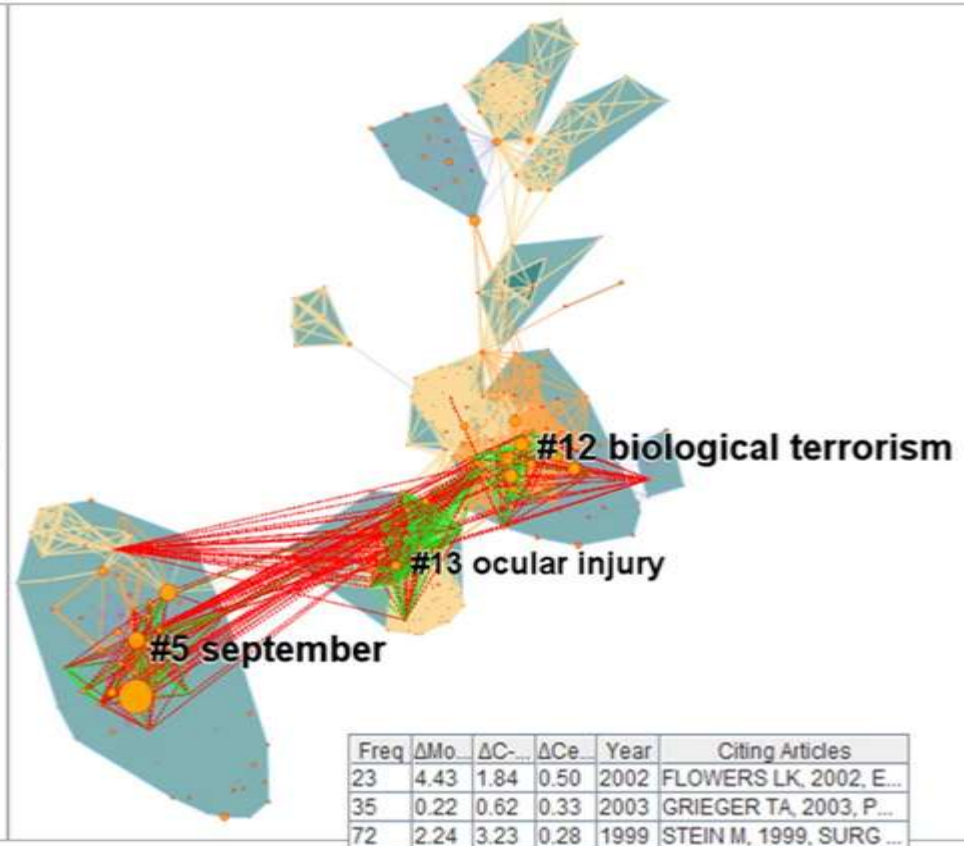
Data Source: Complex Network Analysis (1996-2004), top100 records per time slice, 2-year sliding window										
Unit of Analysis	Reference		Keyword		Noun Phrase		Author		Journal	
Relation	Co-citation		Co-occurrence		Co-occurrence		Co-citation		Co-citation	
Offset (Exposure)	$\log_2$(Year)		$\log_2$(Year)		$\log_2$(Year)		$\log_2$(Year)		$\log_2$(Year)	
Number of Citing Articles	3,515		3,072		3,254		3,271		3,271	
Global Citations	**Incidence Rate Ratios (IRRs) in NB Models**									
Coauthors	**1.306**	**.000**	**1.298**	**.000**	**1.326**	**.000**	**1.359**	**.000**	**1.350**	**.000**
Modularity Change Rate	**1.083**	**.025**	1.038	.086	1.047	.305	1.055	.276	1.060	.180
Weighted Cluster Linkage	**3.160**	**.000**	0.205	.095	**1.33×10^{8}**	**.000**	**2.879**	**.000**	**1.204**	**.049**
Centrality Divergence	0.343	.184	**3.679**	**.023**	1.534	.665	**23.400**	**.000**	**7.620**	**.000**
Number of References	**1.013**	**.000**	**1.013**	**.000**	**1.013**	**.000**	**1.012**	**.000**	**1.012**	**.000**
Number of Pages	**0.970**	**.000**	**0.971**	**.000**	**0.971**	**.000**	**0.973**	**.000**	**0.972**	**.000**
Dispersion Parameter (θ)	0.5284		0.5258		0.5150		0.5282		0.5375	
2 x log-likelihood	-31,771		-28,331		-29,491		-29,506		-29,613	
Akaike's Information Criterion (AIC)	31,787		28,347		29,508		29,522		29,629	

Sources: Chen, C. (2012) Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, **63**(3), 431-449.

Examples: Scientific Literature

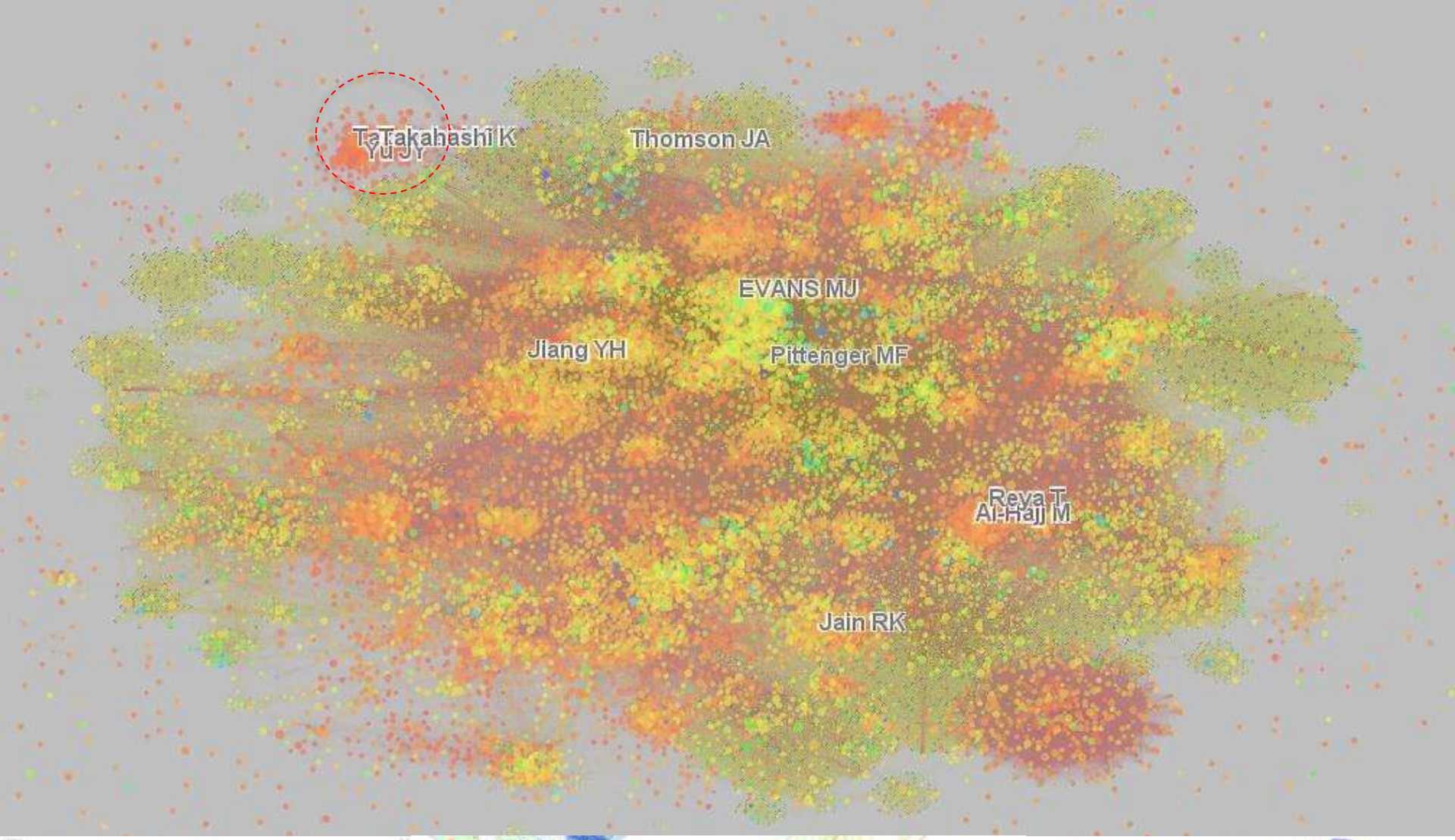
Terrorism

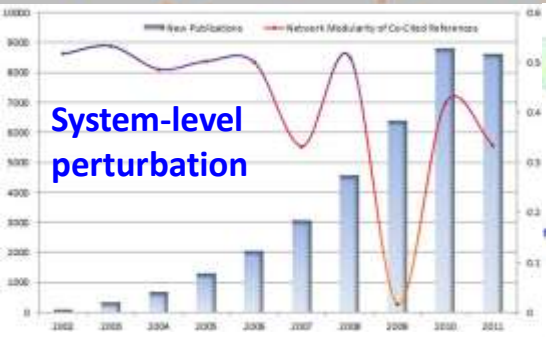




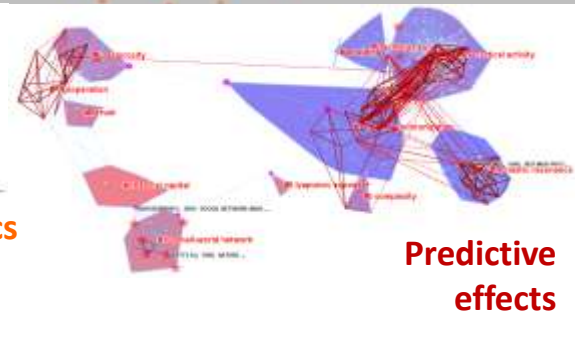
Freq	ΔMo...	ΔC-...	ΔCe...	Year	Citing Articles
23	4.43	1.84	0.50	2002	FLOWERS LK, 2002, E...
35	0.22	0.62	0.33	2003	GRIEGER TA, 2003, P...
72	2.24	3.23	0.28	1999	STEIN M, 1999, SURG ...
80	0.63	4.14	0.27	2002	AHERN J, 2002, PSYC...
23	-0.20	1.45	0.26	2003	FORD CA, 2003, ARCH...

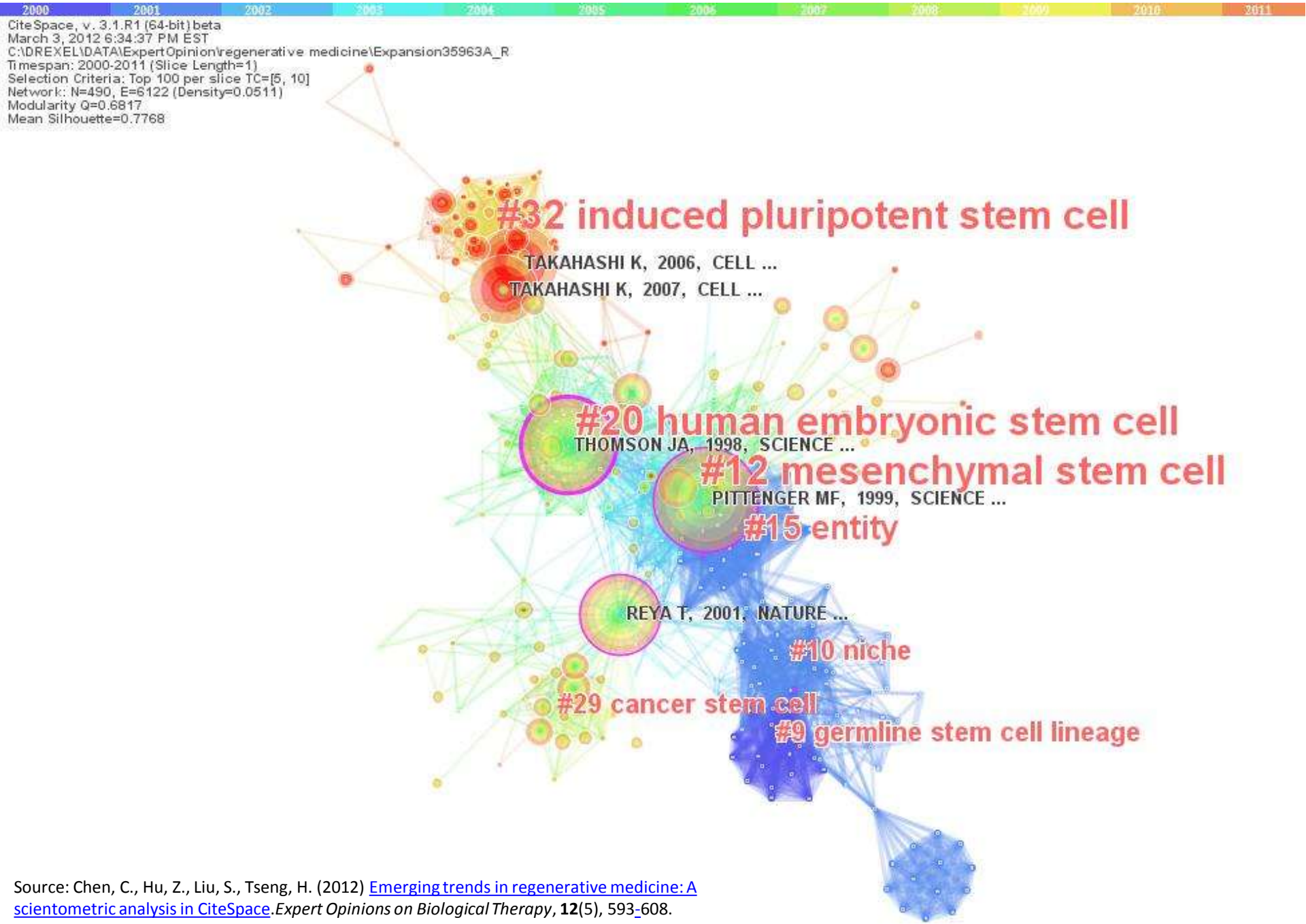
Sources: Chen, C. (2012) Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, **63**(3), 431-449.


Takahashi K
Yu JY
Thomson JA
EVANS MJ
Jiang YH
Pittenger MF
Reya T
Al-Hajj M
Jain RK

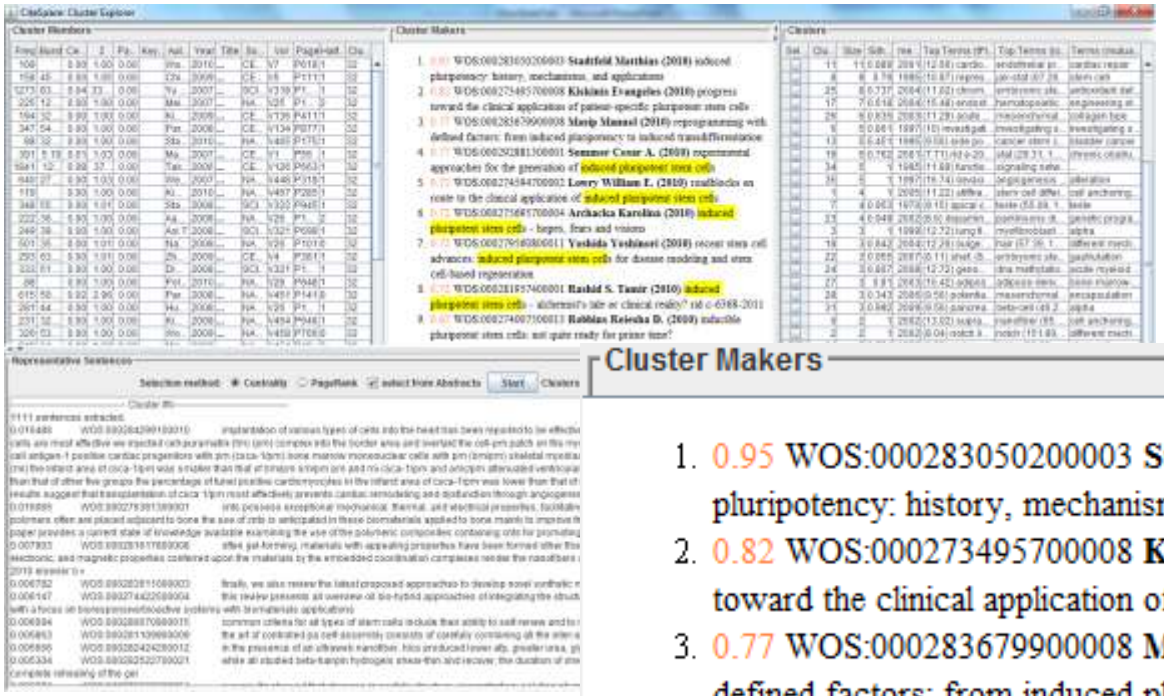

System-level perturbation

#11 mesenchymal stem cell
#15 cancer stem cell
#19 combination
#12 human embryonic stem cell
#7 induced pluripotent stem cell
methylation
Situation awareness
Hot topics


Predictive effects



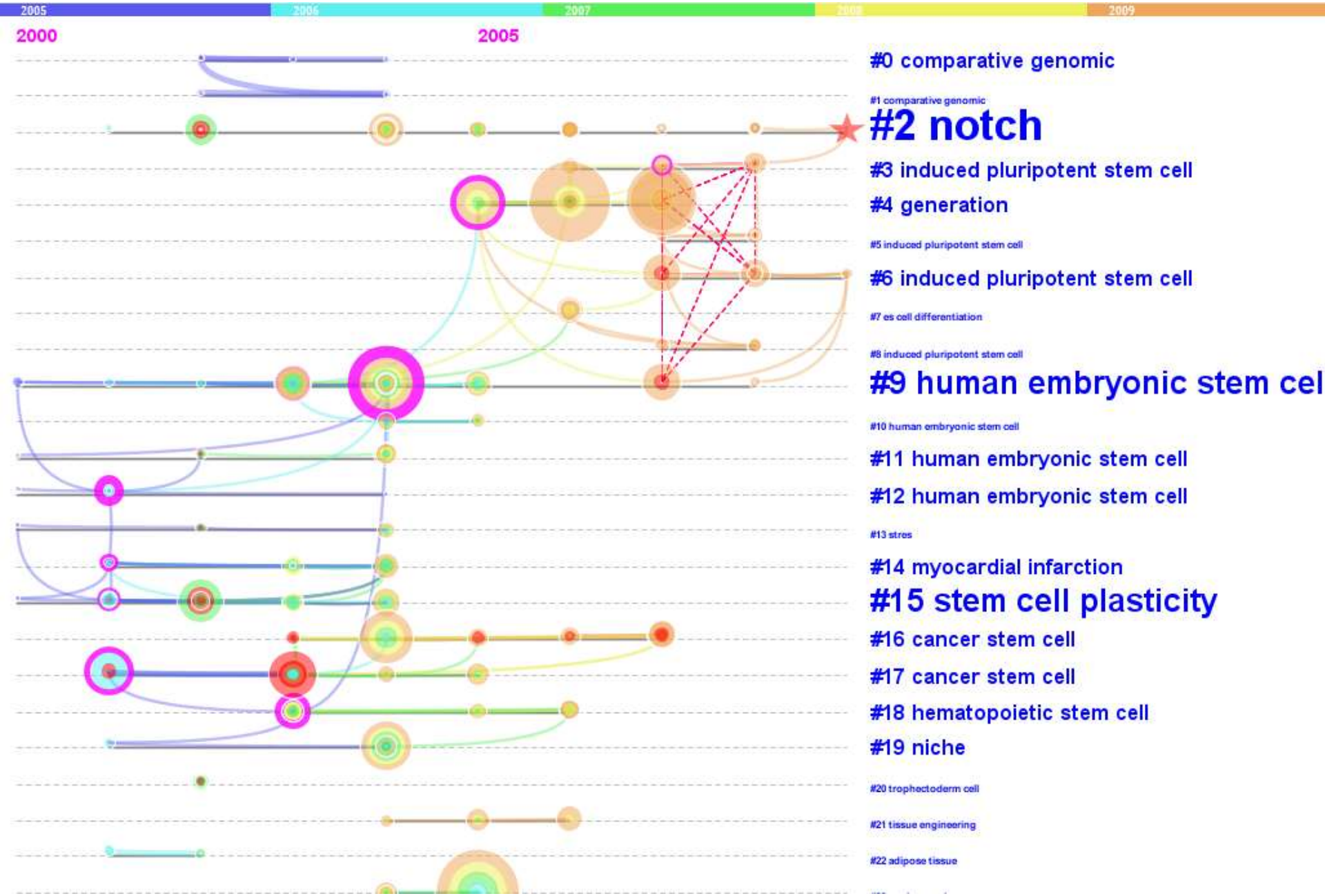
Source: Chen, C., Hu, Z., Liu, S., Tseng, H. (2012) Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinions on Biological Therapy*, **12**(5), 593-608.



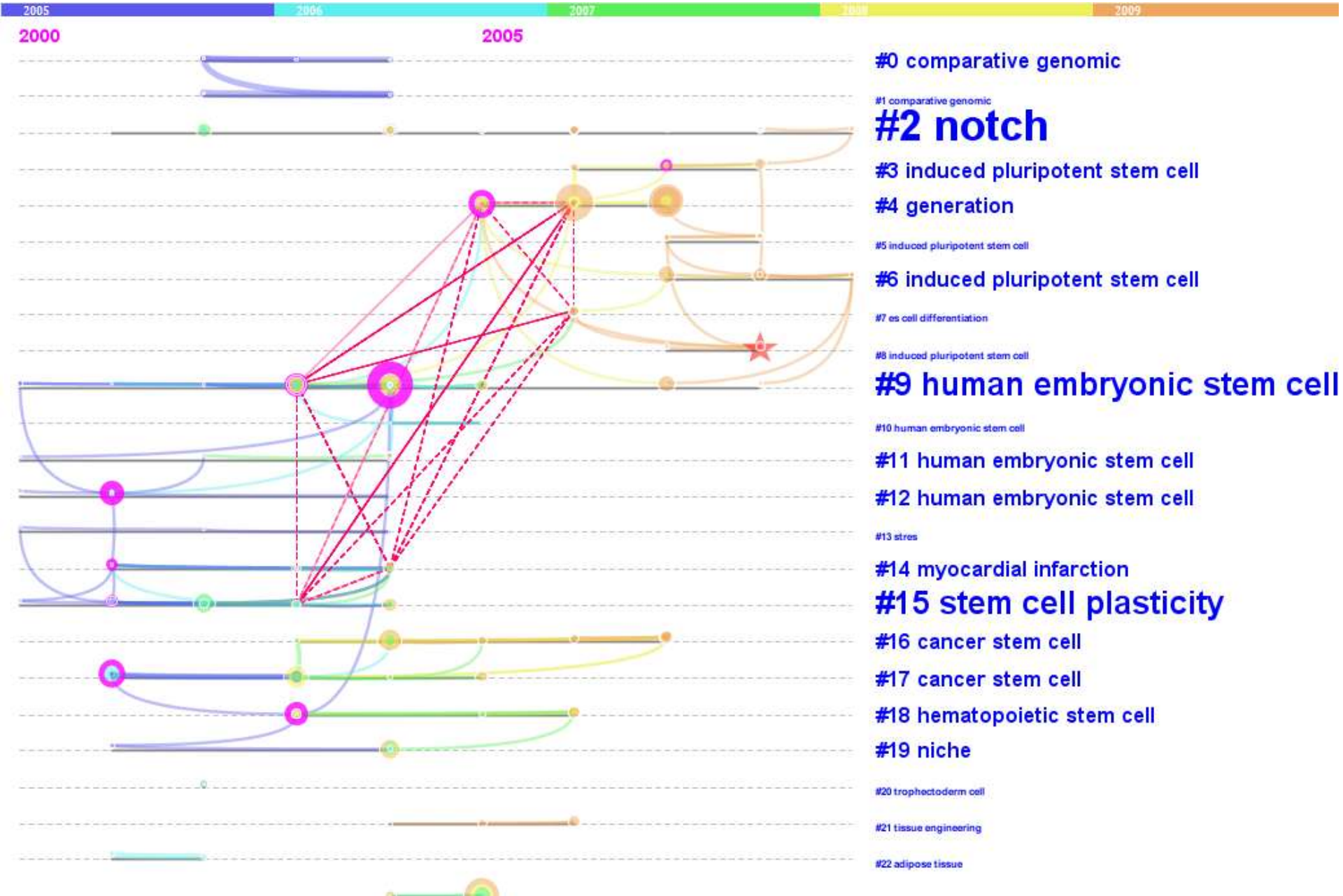
Cluster Makers

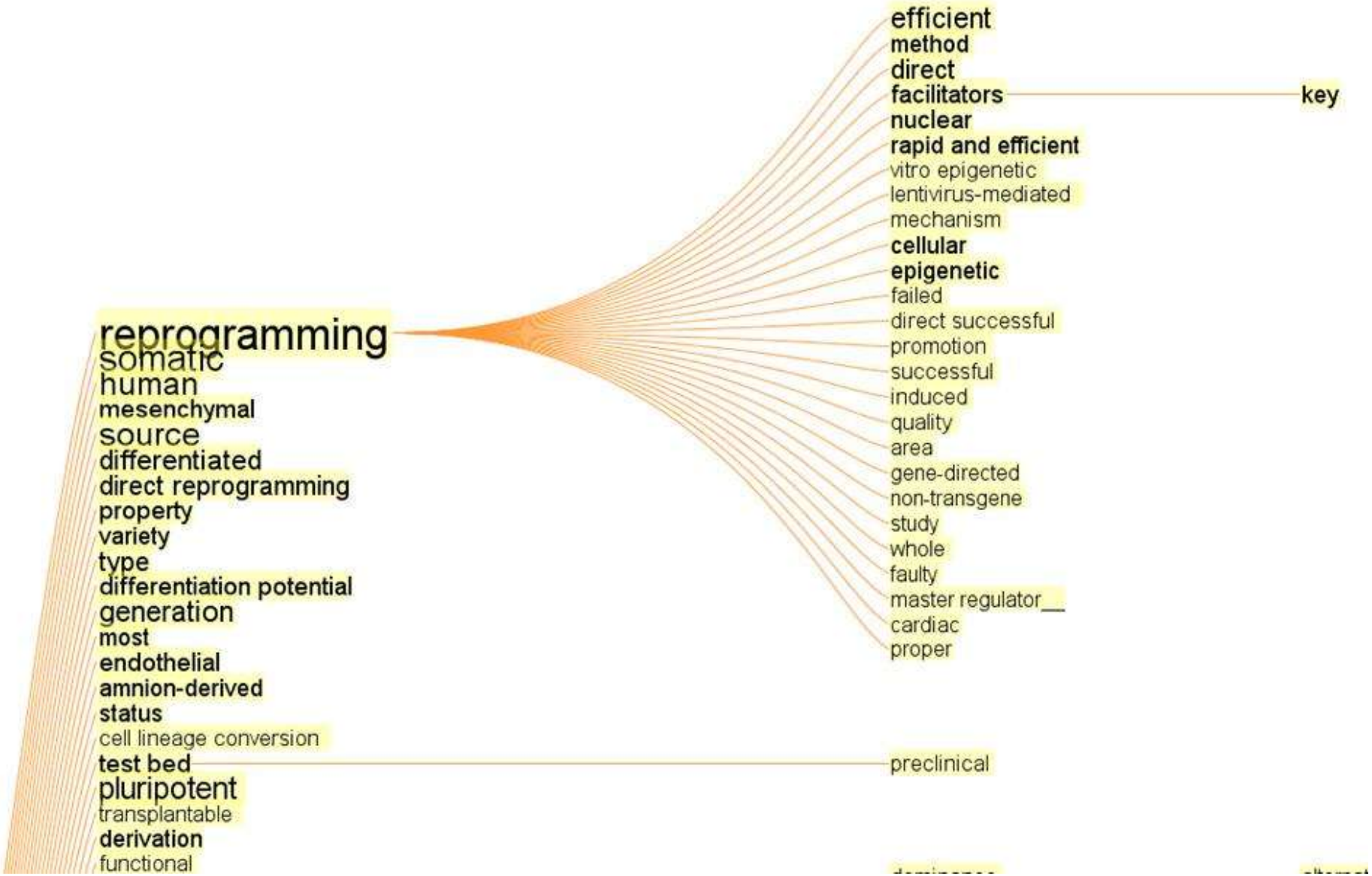
1. 0.95 WOS:000283050200003 **Stadtfeld Matthias (2010)** induced pluripotency: history, mechanisms, and applications
2. 0.82 WOS:000273495700008 **Kiskinis Evangelos (2010)** progress toward the clinical application of patient-specific pluripotent stem cells
3. 0.77 WOS:000283679900008 **Masip Manuel (2010)** reprogramming with defined factors: from induced pluripotency to induced transdifferentiation
4. 0.77 WOS:000292881300001 **Sommer Cesar A. (2010)** experimental approaches for the generation of induced pluripotent stem cells
5. 0.72 WOS:000274594700002 **Lowry William E. (2010)** roadblocks en route to the clinical application of induced pluripotent stem cells
6. 0.72 WOS:000275695700004 **Archacka Karolina (2010)** induced pluripotent stem cells - hopes, fears and visions
7. 0.72 WOS:000279560800011 **Yoshida Yoshinori (2010)** recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration
8. 0.72 WOS:000281957400001 **Rashid S. Tamir (2010)** induced pluripotent stem cells - alchemist's tale or clinical reality? rid c-6368-2011
9. 0.67 WOS:000274007500013 **Robbins Reiesha D. (2010)** inducible pluripotent stem cells: not quite ready for prime time?

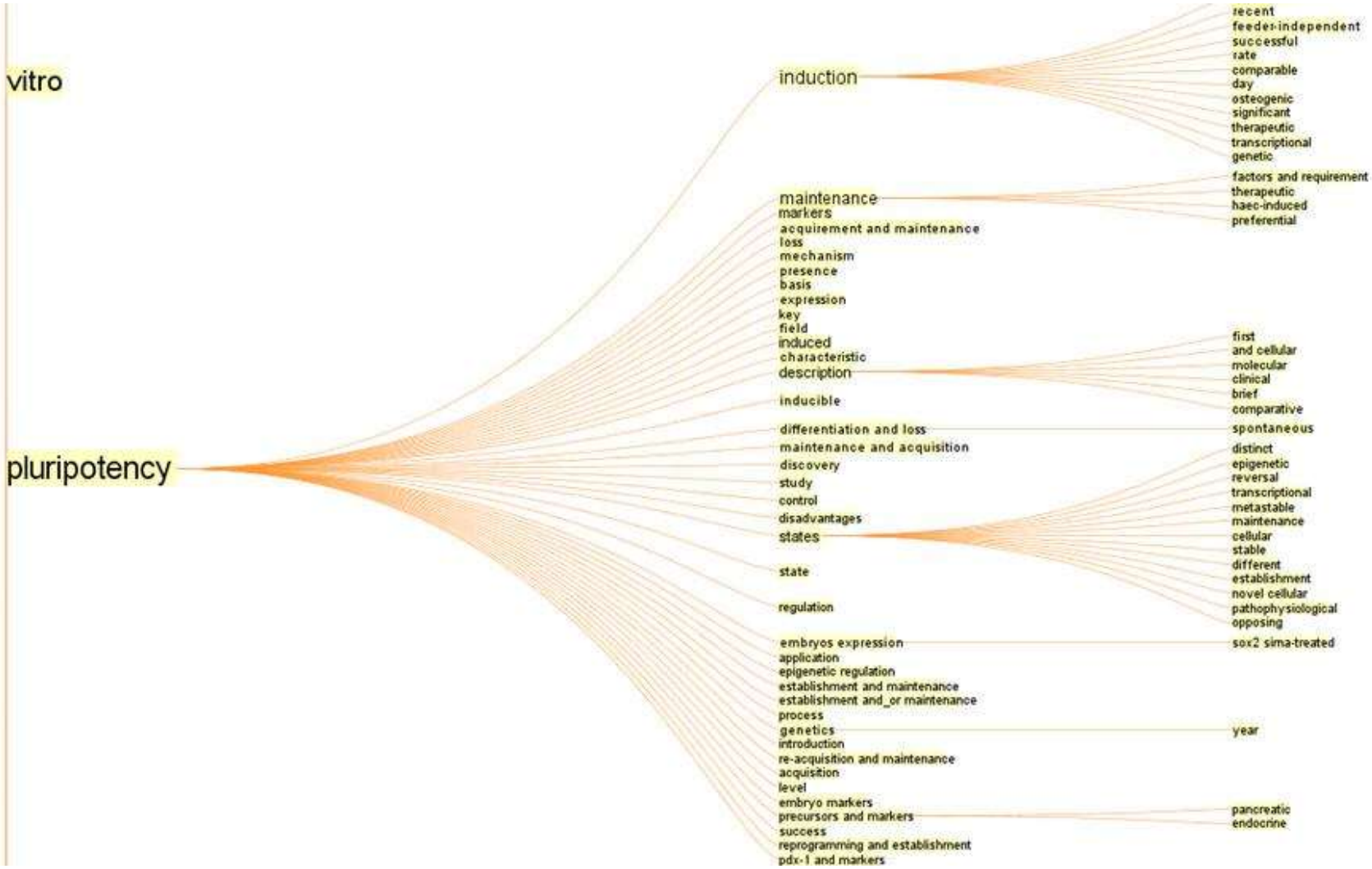
KAJI KEISUKE, 2009, NATURE, V458, P771, 10.1038/nature07864 Cited 321.

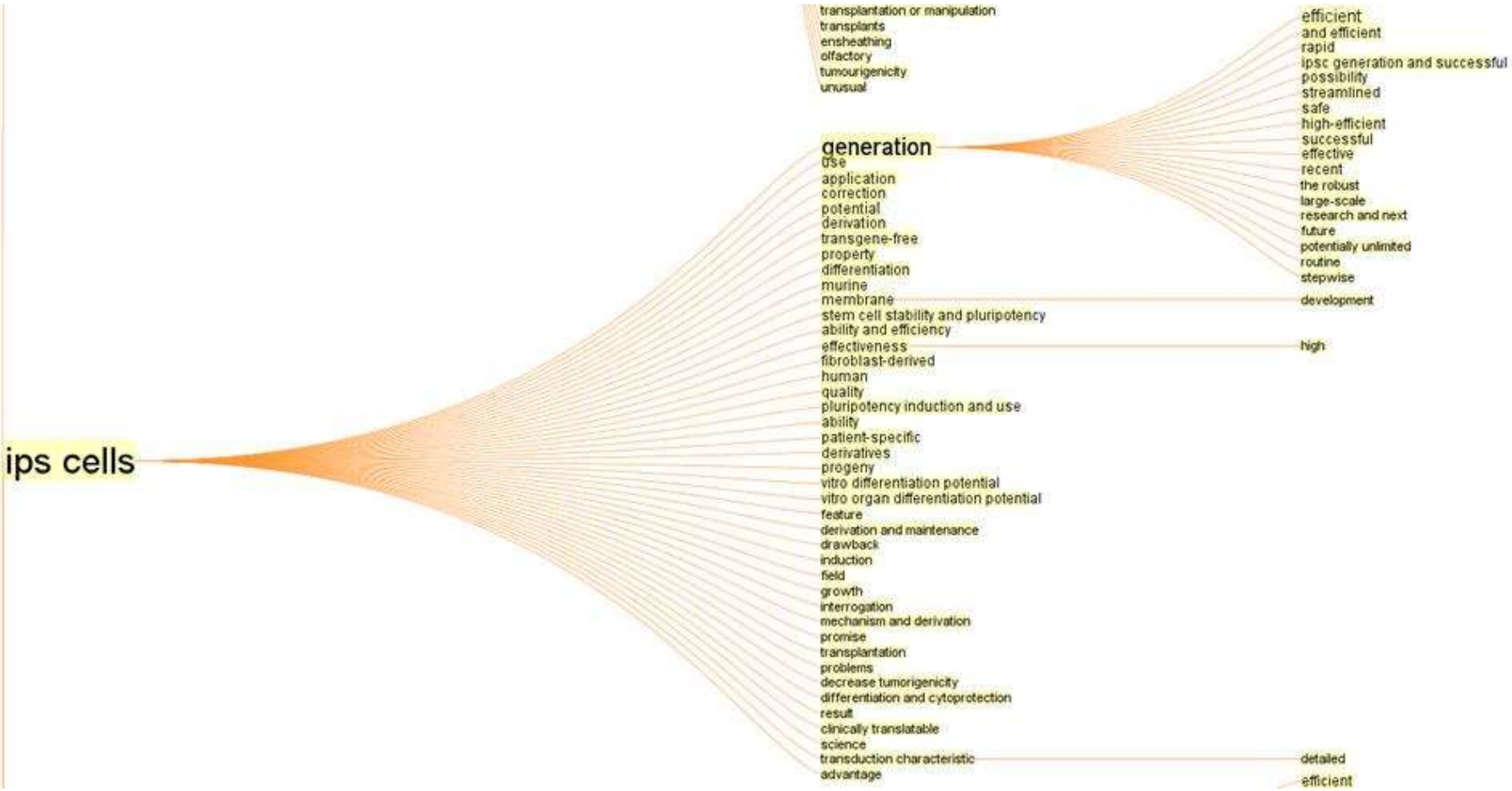


JAENISCH RUDOLF, 2008, CELL, V132, P567, 10.1016/j.cell.2008.01.015 Cited 379 times



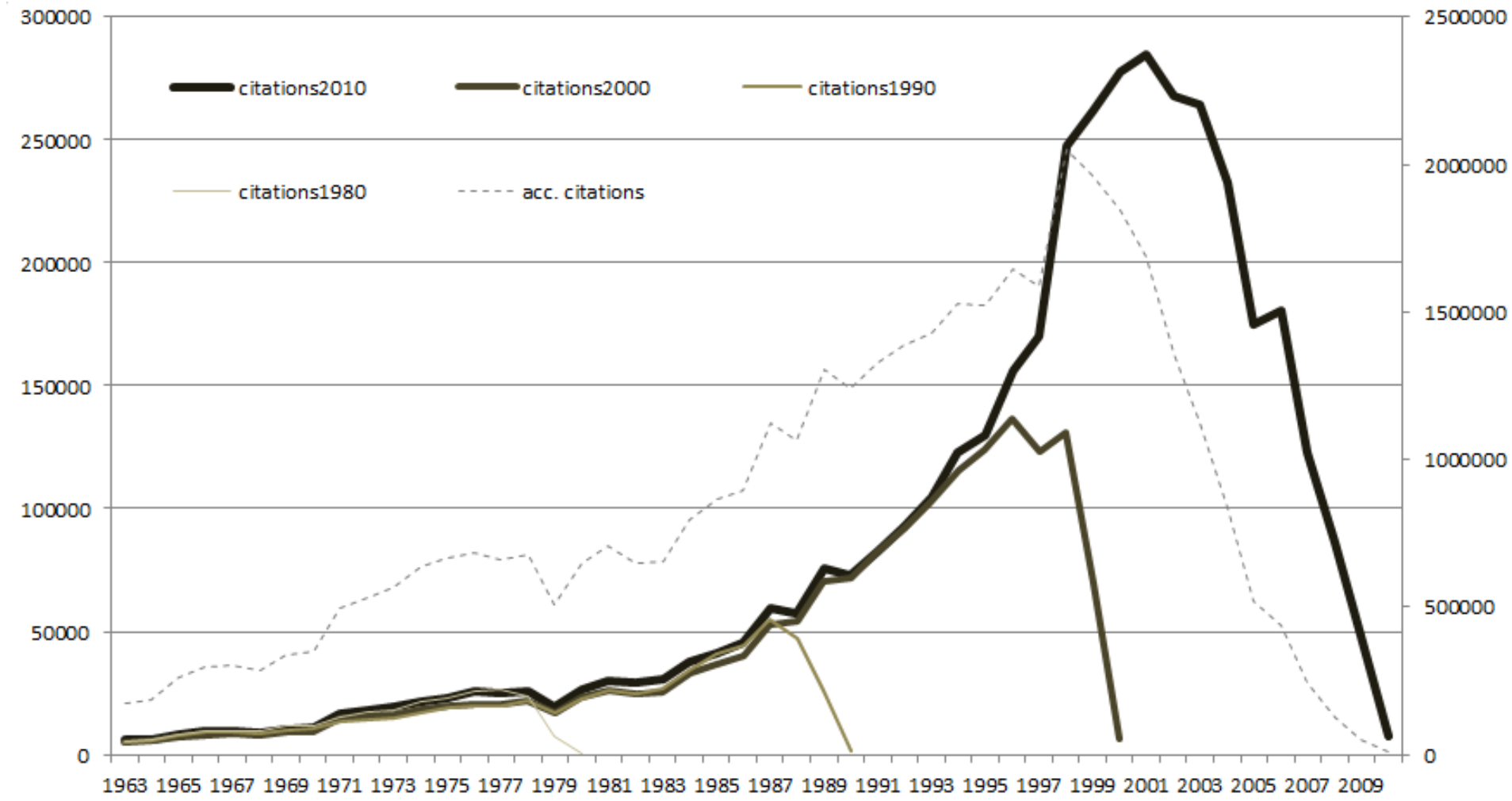

reprogramming
somatic
human
mesenchymal
source
differentiated
direct reprogramming
property
variety
type
differentiation potential
generation
most
endothelial
amnion-derived
status
cell lineage conversion
test bed
pluripotent
transplantable
derivation
functional
efficient
method
direct
facilitators
key
nuclear
rapid and efficient
vitro epigenetic
lentivirus-mediated
mechanism
cellular
epigenetic
failed
direct successful
promotion
successful
induced
quality
area
gene-directed
non-transgene
study
whole
faulty
master regulator
cardiac
proper
preclinical


vitro
pluripotency
induction
recent
feeder-independent
successful
rate
comparable
day
osteogenic
significant
therapeutic
transcriptional
genetic
maintenance
factors and requirement
therapeutic
haec-induced
preferential
markers
acquirement and maintenance
loss
mechanism
presence
basis
expression
key
field
induced
characteristic
description
first
and cellular
molecular
clinical
brief
comparative
inducible
differentiation and loss
spontaneous
maintenance and acquisition
discovery
study
control
disadvantages
states
distinct
epigenetic
reversal
transcriptional
metastable
maintenance
cellular
stable
different
establishment
novel cellular
pathophysiological
opposing
state
regulation
embryos expression
sox2 sima-treated
application
epigenetic regulation
establishment and maintenance
establishment and_or maintenance
process
genetics
year
introduction
re-acquisition and maintenance
acquisition
level
embryo markers
precursors and markers
pancreatic
endocrine
success
reprogramming and establishment
pdx-1 and markers


ips cells
transplantation or manipulation
transplants
ensheathing
olfactory
tumourigenicity
unusual
generation
use
application
correction
potential
derivation
transgene-free
property
differentiation
murine
membrane
stem cell stability and pluripotency
ability and efficiency
effectiveness
fibroblast-derived
human
quality
pluripotency induction and use
ability
patient-specific
derivatives
progeny
vitro differentiation potential
vitro organ differentiation potential
feature
derivation and maintenance
drawback
induction
field
growth
interrogation
mechanism and derivation
promise
transplantation
problems
decrease tumorigenicity
differentiation and cytoprotection
result
clinically translatable
science
transduction characteristic
advantage
efficient
and efficient
rapid
ipsc generation and successful
possibility
streamlined
safe
high-efficient
successful
effective
recent
the robust
large-scale
research and next
future
potentially unlimited
routine
stepwise
development
high
detailed
efficient

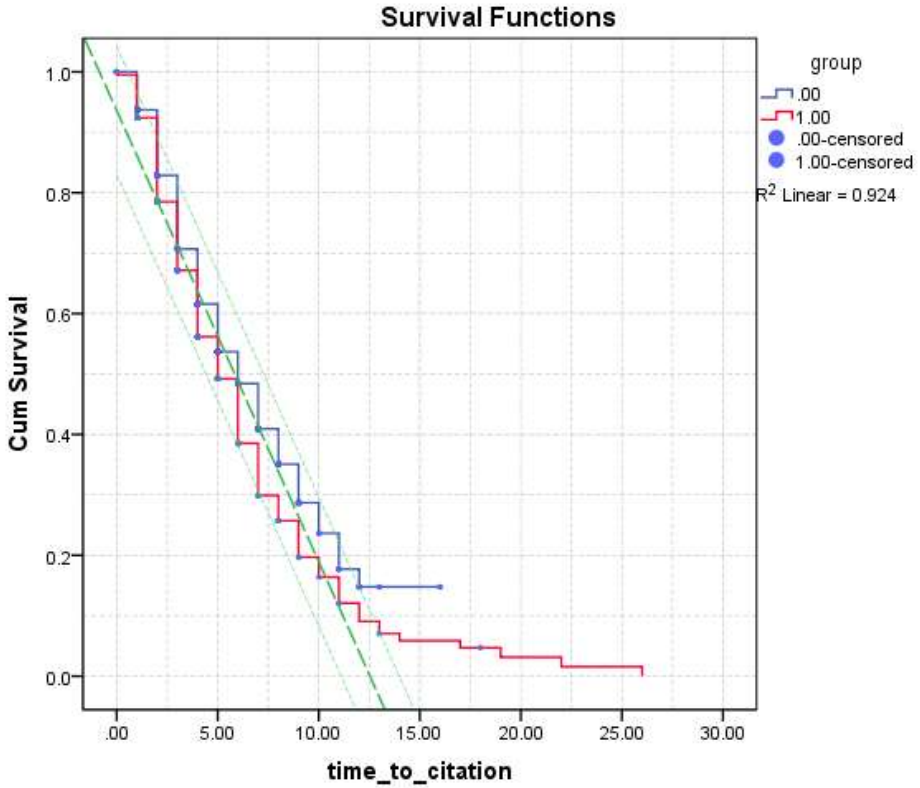
Examples: Patents

Time Taken to Citation



NCI Patents: Time to Citation

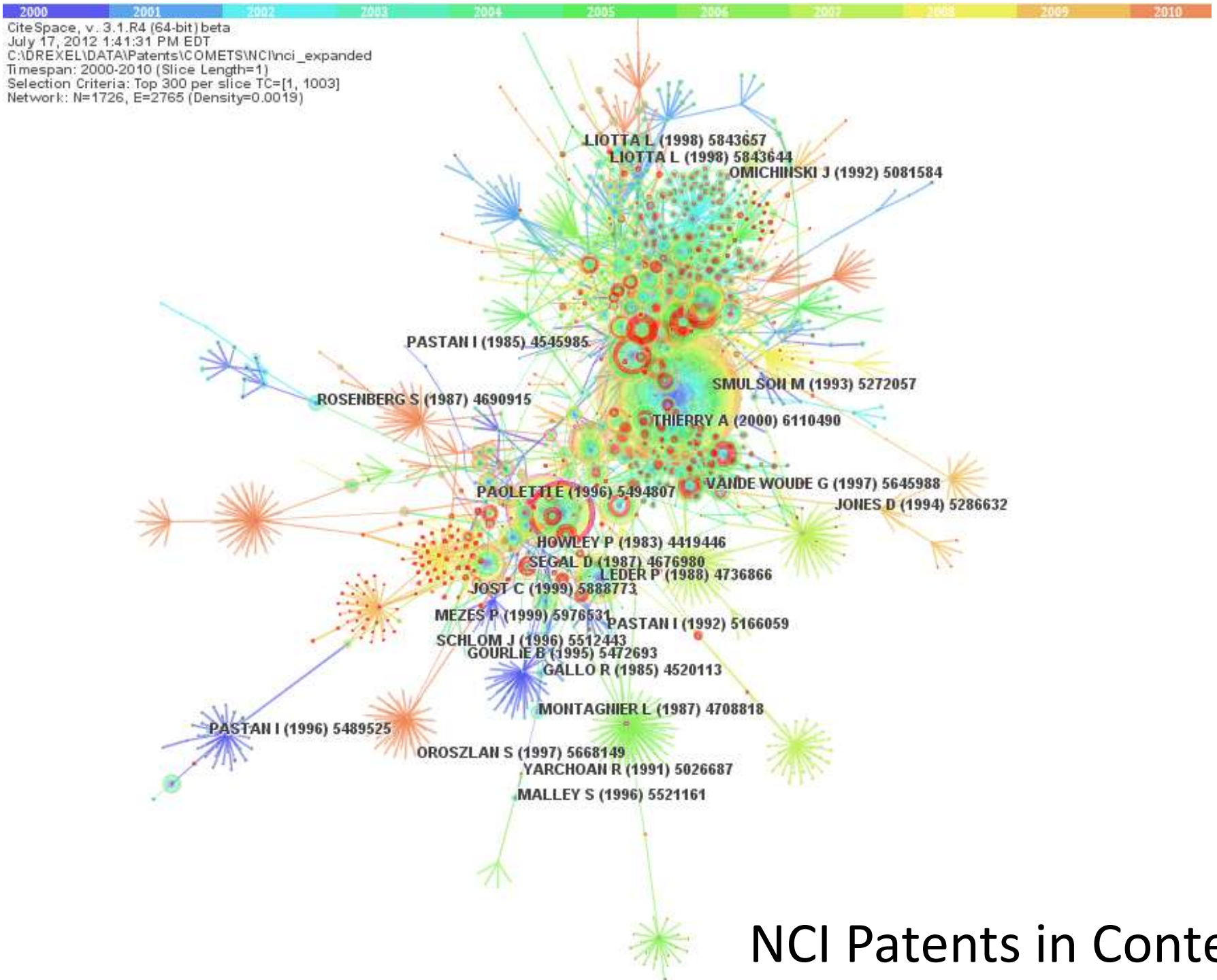
blue (0) – active; red (1) - inactive



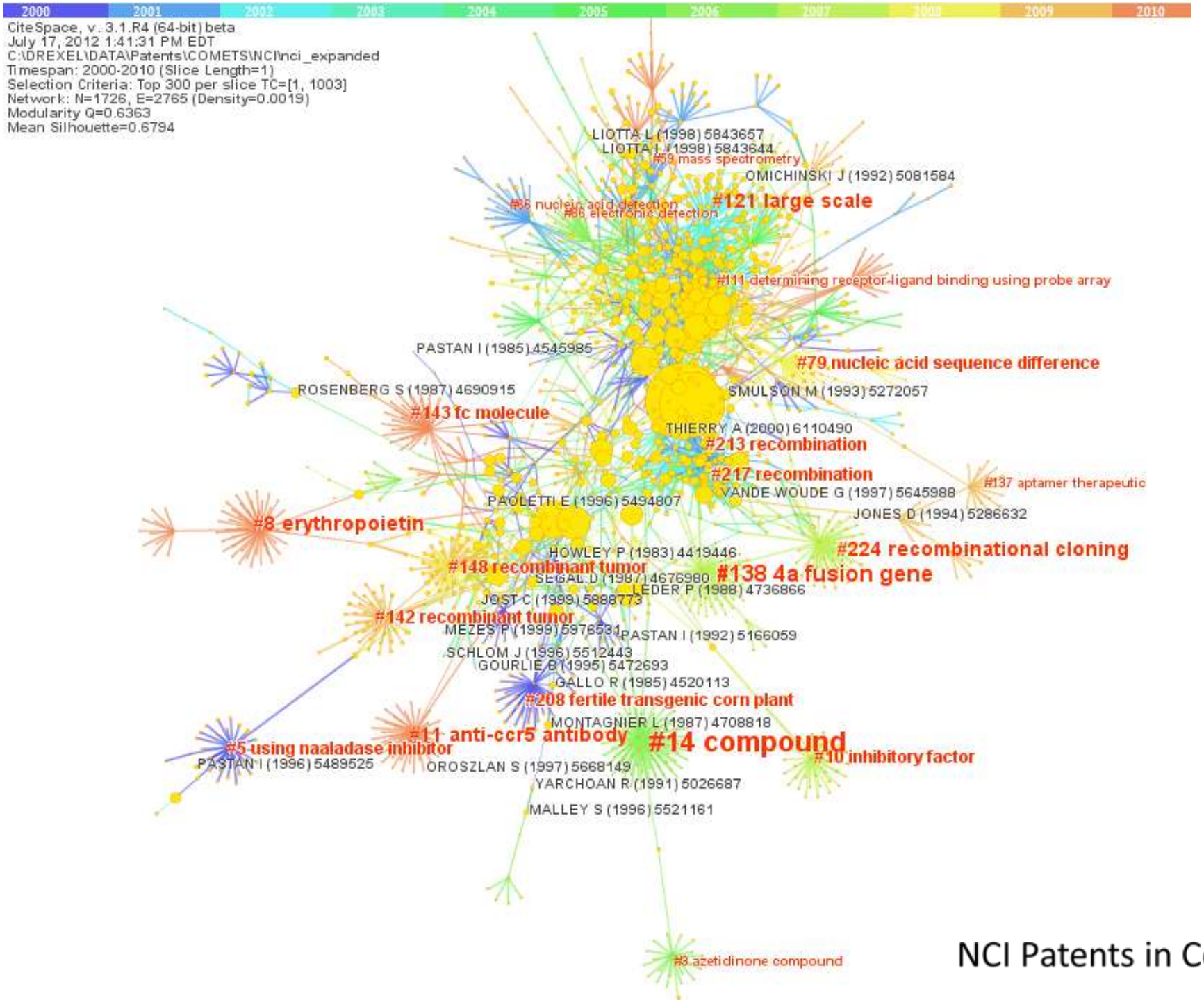
Means and Medians for Survival Time

group	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
.00	7.161	.350	6.475	7.846	6.000	.431	5.155	6.845
1.00	6.440	.371	5.713	7.167	5.000	.344	4.326	5.674
Overall	7.029	.337	6.369	7.690	6.000	.289	5.433	6.567

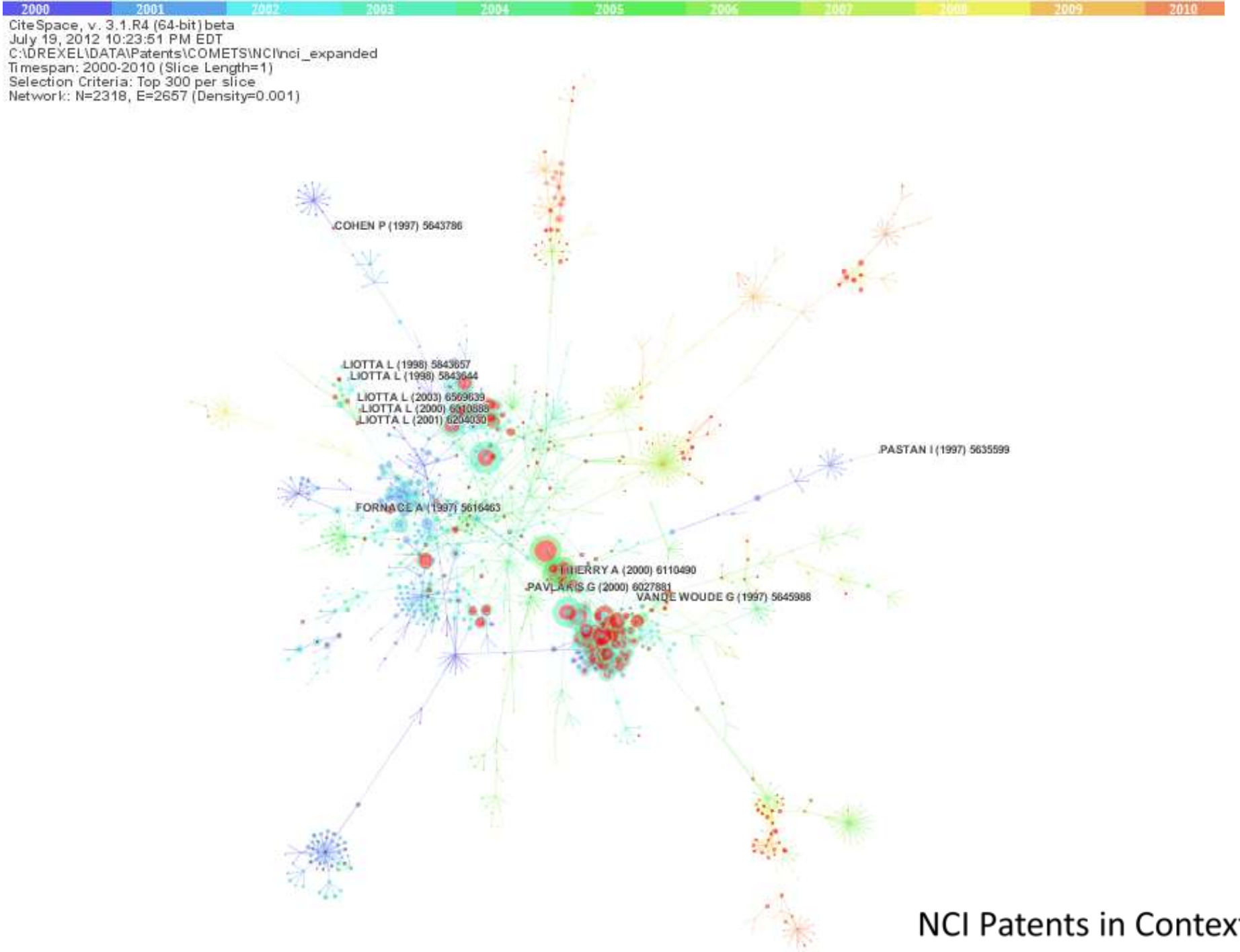
a. Estimation is limited to the largest survival time if it is censored.



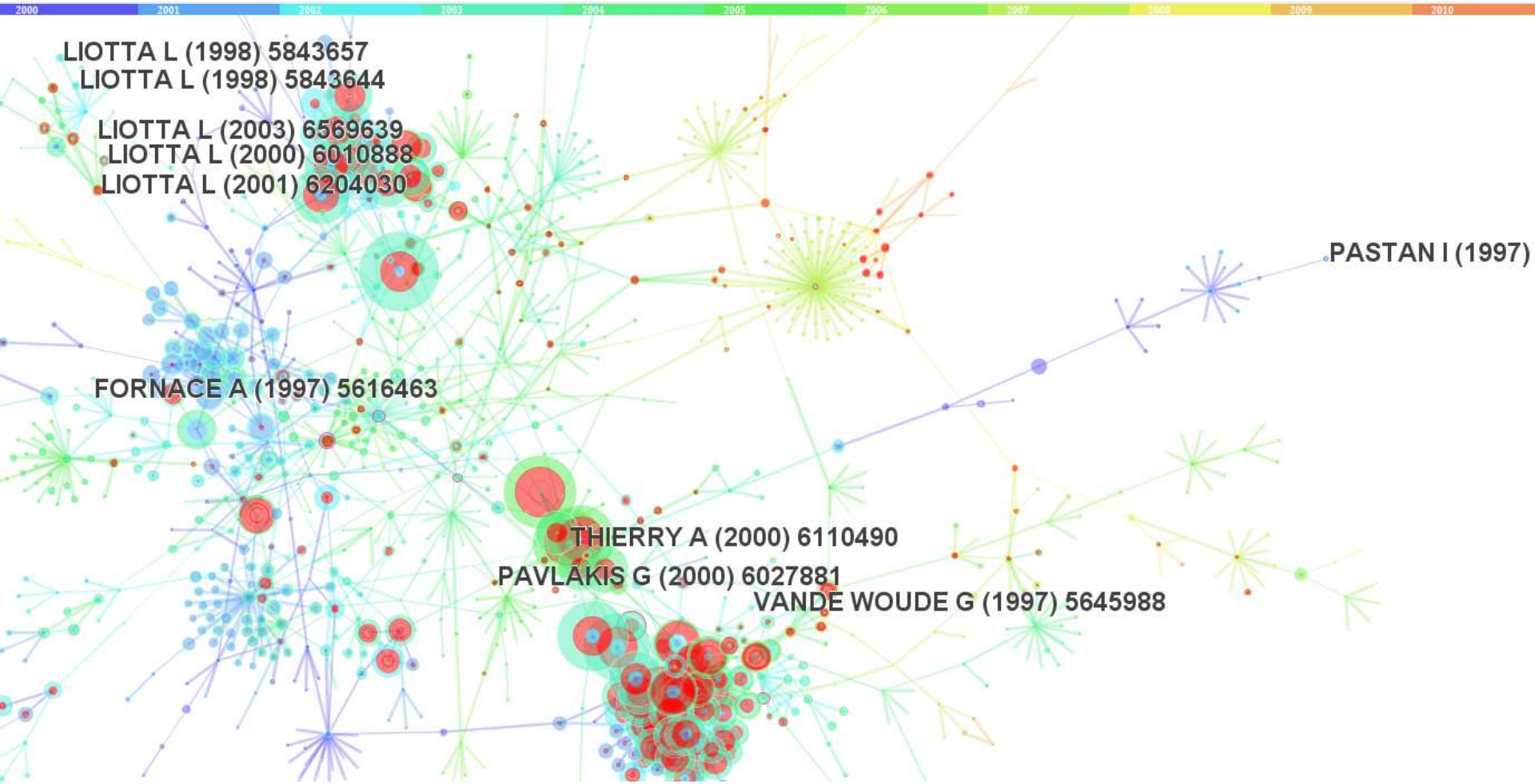
NCI Patents in Context



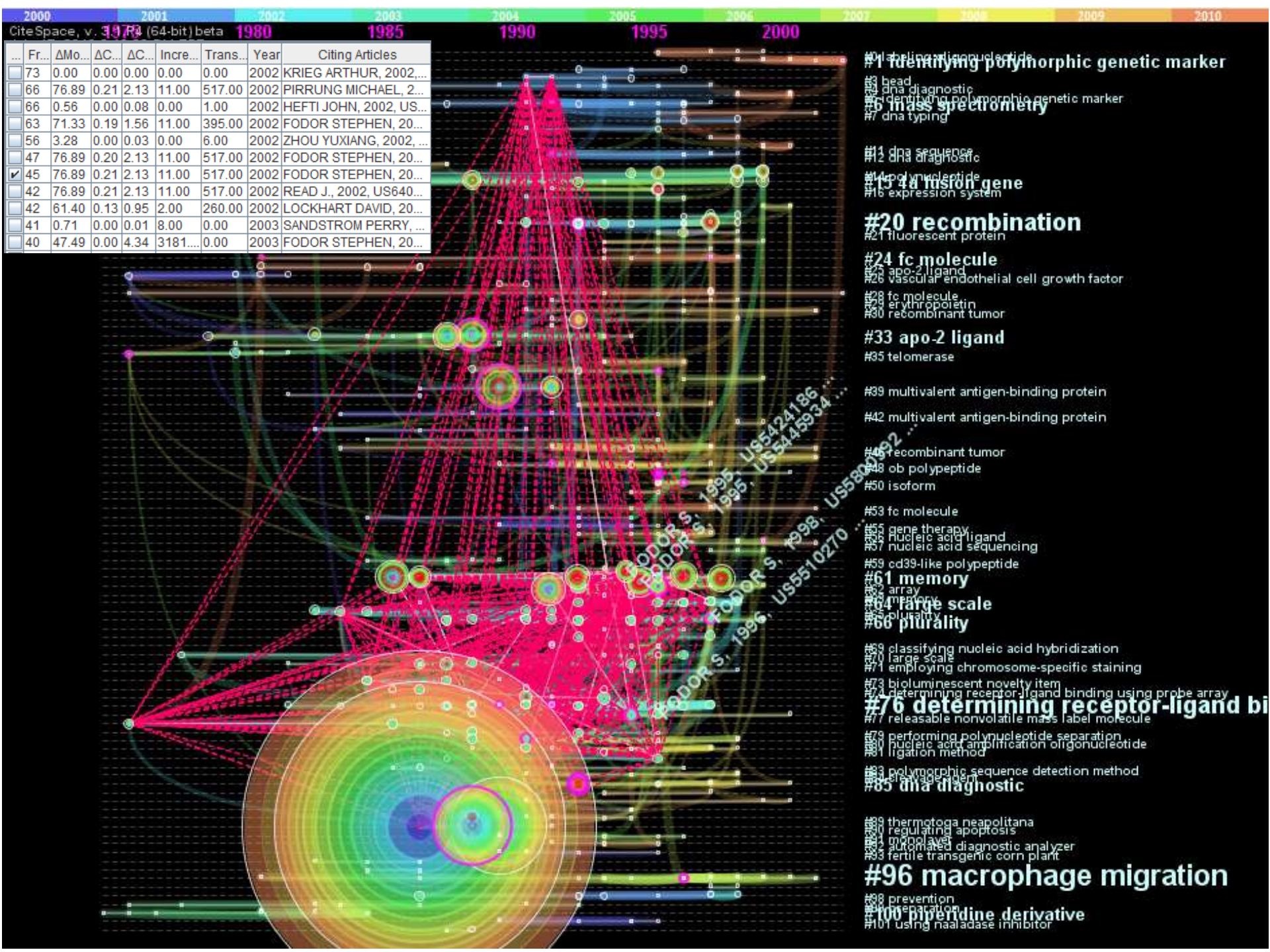
NCI Patents in Context



NCI Patents in Context



NCI Patents in Context


2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
CiteSpace, v. 3.7.R4 (64-bit) beta
1980
1985
1990
1995
2000
... Fr... ΔMo... ΔC... ΔC... Incre... Trans... Year Citing Articles
73 0.00 0.00 0.00 0.00 0.00 2002 KRIEG ARTHUR, 2002,...
66 76.89 0.21 2.13 11.00 517.00 2002 PIRRUNG MICHAEL, 2...
66 0.56 0.00 0.08 0.00 1.00 2002 HEFTI JOHN, 2002, US...
63 71.33 0.19 1.56 11.00 395.00 2002 FODOR STEPHEN, 20...
56 3.28 0.00 0.03 0.00 6.00 2002 ZHOU YUXIANG, 2002, ...
47 76.89 0.20 2.13 11.00 517.00 2002 FODOR STEPHEN, 20...
45 76.89 0.21 2.13 11.00 517.00 2002 FODOR STEPHEN, 20...
42 76.89 0.21 2.13 11.00 517.00 2002 READ J., 2002, US640...
42 61.40 0.13 0.95 2.00 260.00 2002 LOCKHART DAVID, 20...
41 0.71 0.00 0.01 8.00 0.00 2003 SANDSTROM PERRY, ...
40 47.49 0.00 4.34 3181.... 0.00 2003 FODOR STEPHEN, 20...
#1 identifying polymorphic genetic marker
#3 bead
#4 dna diagnostic
#5 identifying polymorphic genetic marker
#6 mass spectrometry
#7 dna typing
#11 dna sequence
#12 dna diagnostic
#14 polynucleotide
#15 4a insron gene
#16 expression system
#20 recombination
#21 fluorescent protein
#24 fc molecule
#25 apo-2 ligand
#26 vascular endothelial cell growth factor
#28 fc molecule
#29 erythropoietin
#30 recombinant tumor
#33 apo-2 ligand
#35 telomerase
#39 multivalent antigen-binding protein
#42 multivalent antigen-binding protein
#46 recombinant tumor
#48 ob polypeptide
#50 isoform
#53 fc molecule
#55 gene therapy
#56 nucleic acid ligand
#57 nucleic acid sequencing
#59 cd39-like polypeptide
#61 memory
#62 array
#63 memory
#64 large scale
#65 plurality
#66 plurality
#69 classifying nucleic acid hybridization
#70 large scale
#71 employing chromosome-specific staining
#73 bioluminescent novelty item
#74 determining receptor-ligand binding using probe array
#76 determining receptor-ligand bi
#77 releasable nonvolatile mass label molecule
#79 performing polynucleotide separation
#80 nucleic acid amplification oligonucleotide
#81 ligation method
#83 polymorphic sequence detection method
#84 cleavage agent
#85 dna diagnostic
#89 thermotoga neapolitana
#90 regulating apoptosis
#91 monolayer
#92 automated diagnostic analyzer
#93 fertile transgenic corn plant
#96 macrophage migration
#98 prevention
#99 separation
#100 piperidine derivative
#101 using naaladase inhibitor
FODOR S, 1995, US5424186 ...
FODOR S, 1995, US5445934 ...
FODOR S, 1998, US5800992 ...
FODOR S, 1996, US5510270 ...

Predictive Effects of Structural Variation Variables on Patent Citations

Top 100 per slice, MST, Look Back 5 Years

mean(globalCites)=3.0077; variance(globalCites)=43.04451

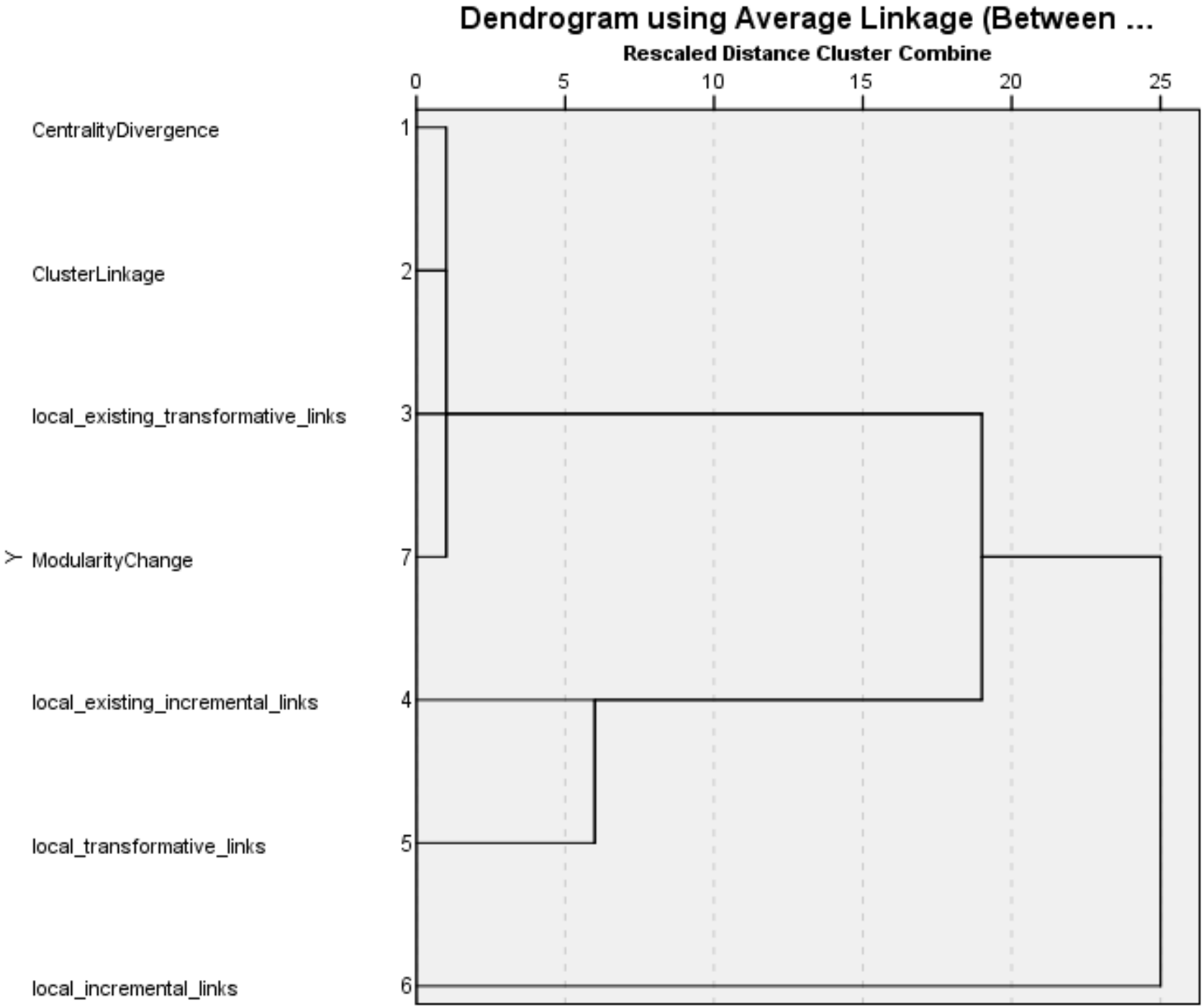
N=2,065; Theta=0.3504 AIC(zinb)=8,146	Count model coefficients			Zero-inflation model		
Variables	IRR	z	p	IRR	z	P
Intercept	0.0001	-142.301	0.0000	0.1374	-4.146	0.0000
Existing transformative links	1.1604	1.953	0.0508	0.6295	-0.290	0.7716
Existing incremental links	**1.0216**	2.595	0.0095	**0.0834**	-2.371	0.0177
Novel transformative links	0.9998	NA	NA	**20.1540**	2.103	0.0355
Novel incremental links	**1.0047**	2.619	0.0088	0.4987	-1.771	0.0765
Modularity	**1.0351**	2.107	0.0351	0.0061	-1.818	0.0690
Cluster linkage	0.1379	NA	NA	0.0000	-1.938	0.0526
Centrality divergence	**22.2186**	3.399	0.0007	**287.4846**	2.189	0.0286

ZINB is better than NB

(N=2,065, theta 0.2794, AIC=8174.781)

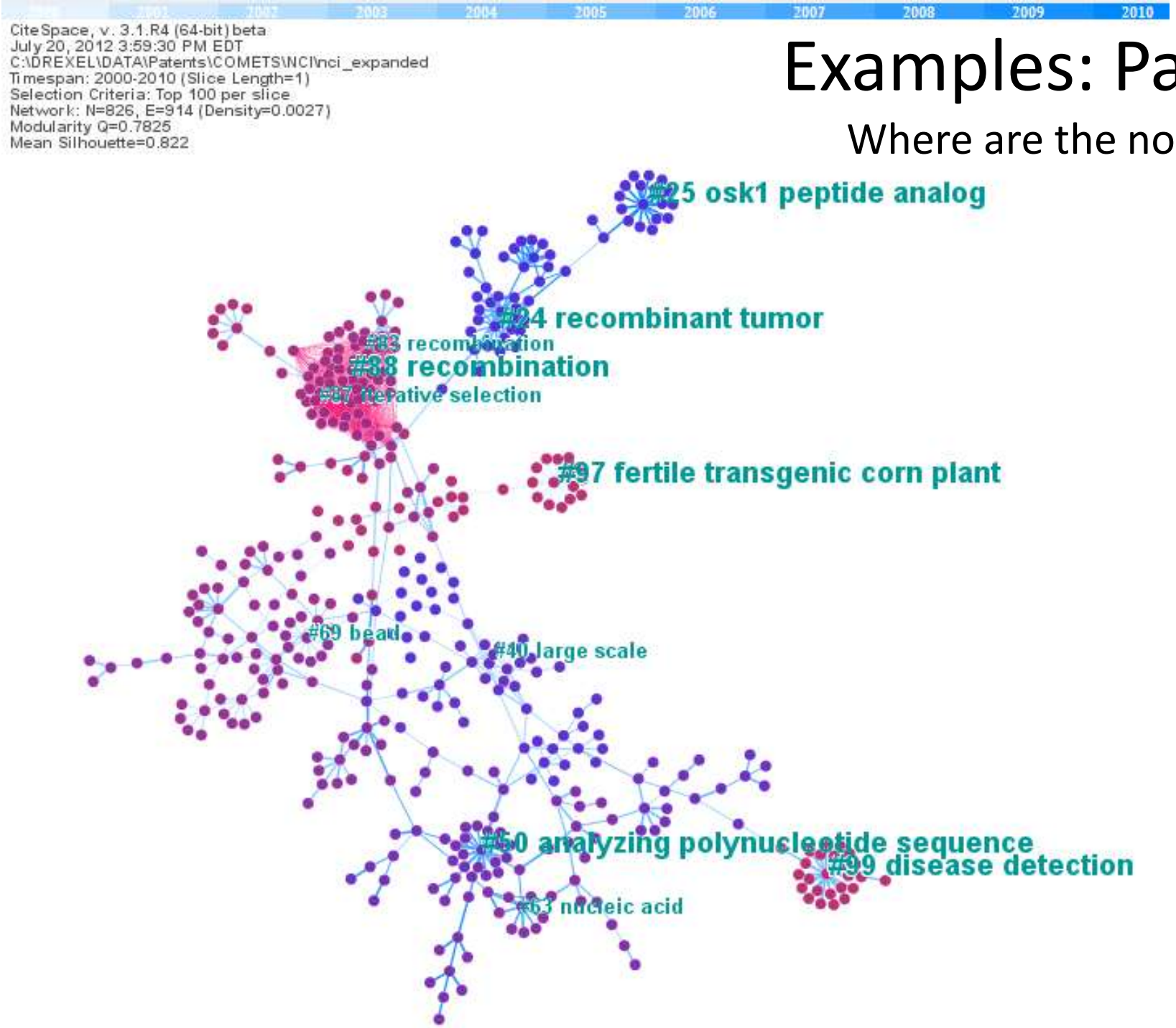
Variables	IRR	Z	p
Intercept	2.664	21.834	0.0000
Local existing transformative links	**1.295**	2.835	0.00458
Local existing incremental links	**1.038**	4.763	0.00000
Local novel transformative links	1.068	1.562	0.11838
Modularity	**0.991**	-0.664	0.50655
Cluster linkage	**0.000**	-2.154	0.03125
Centrality divergence	**3.075**	3.199	0.00138

Vuong Non-Nested Hypothesis Test-Statistic: -3.12993
ZINB > NB, p = 0.00087


Dendrogram using Average Linkage (Between ...
Rescaled Distance Cluster Combine
0
5
10
15
20
25
CentralityDivergence
1
ClusterLinkage
2
local_existing_transformative_links
3
ModularityChange
7
local_existing_incremental_links
4
local_transformative_links
5
local_incremental_links
6

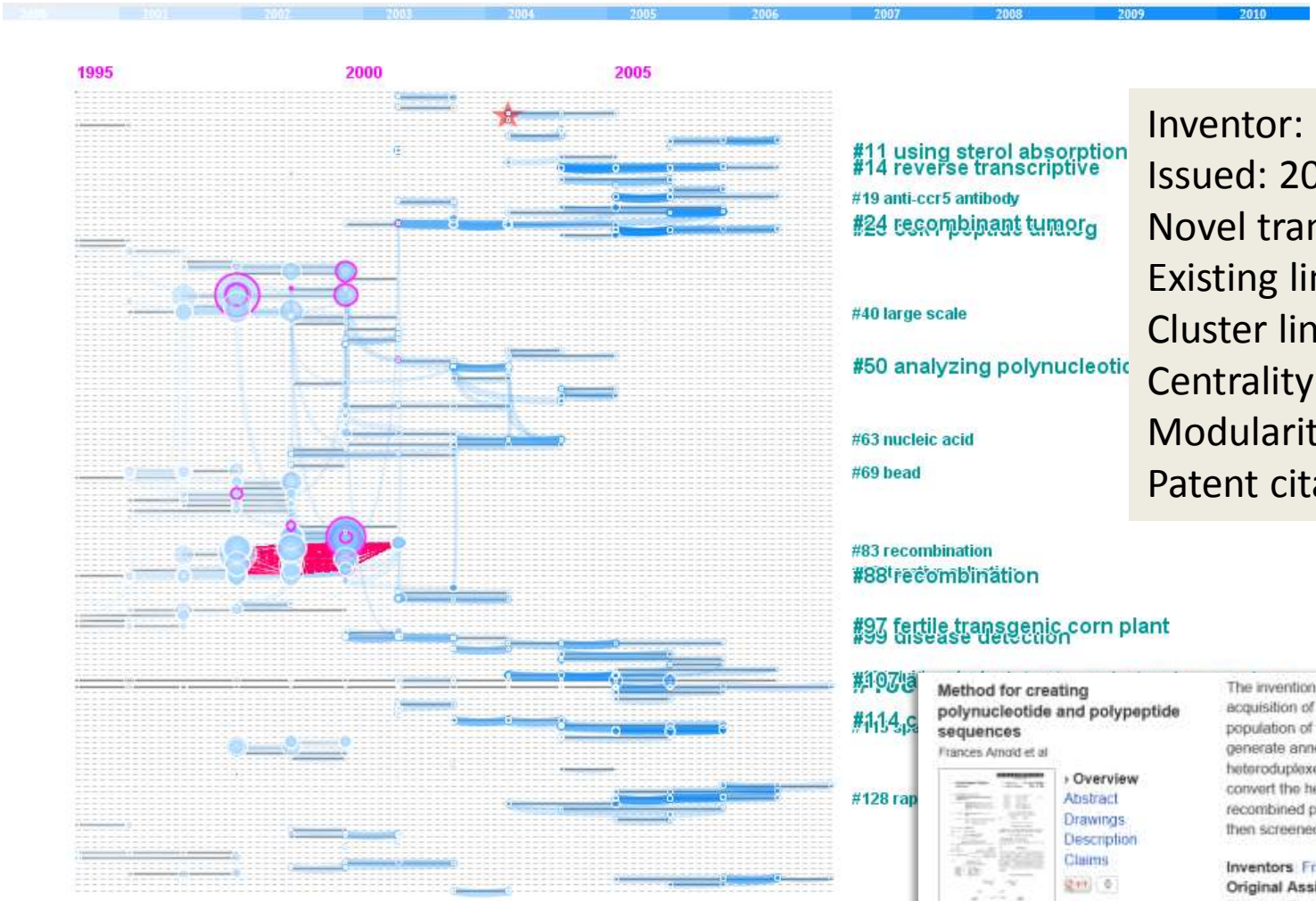
Examples: Patents

Where are the novel links?

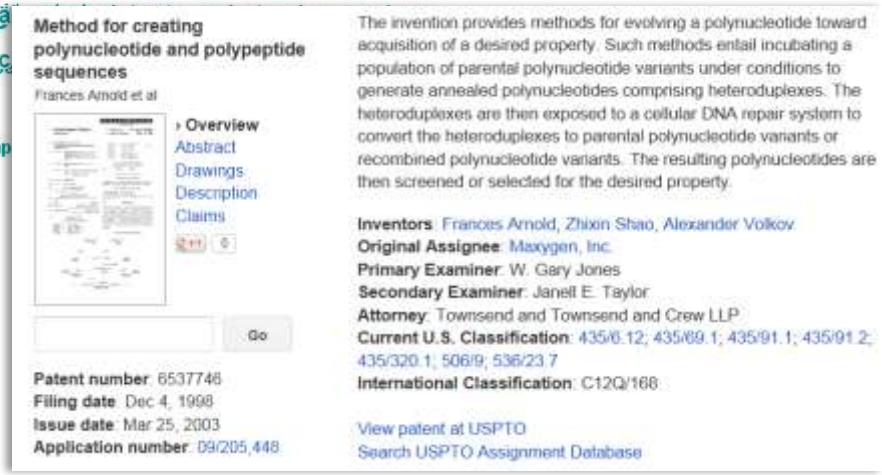


Examples: Patents

US6537746



Inventor: Arnold Frances
Issued: 2003
Novel transformative links: 651
Existing links: 36
Cluster linkage index: 3.50
Centrality divergence: 1.02
Modularity change: 39.75
Patent citations: 22



Method for creating polynucleotide and polypeptide sequences

Frances Arnold et al

› Overview
Abstract
Drawings
Description
Claims

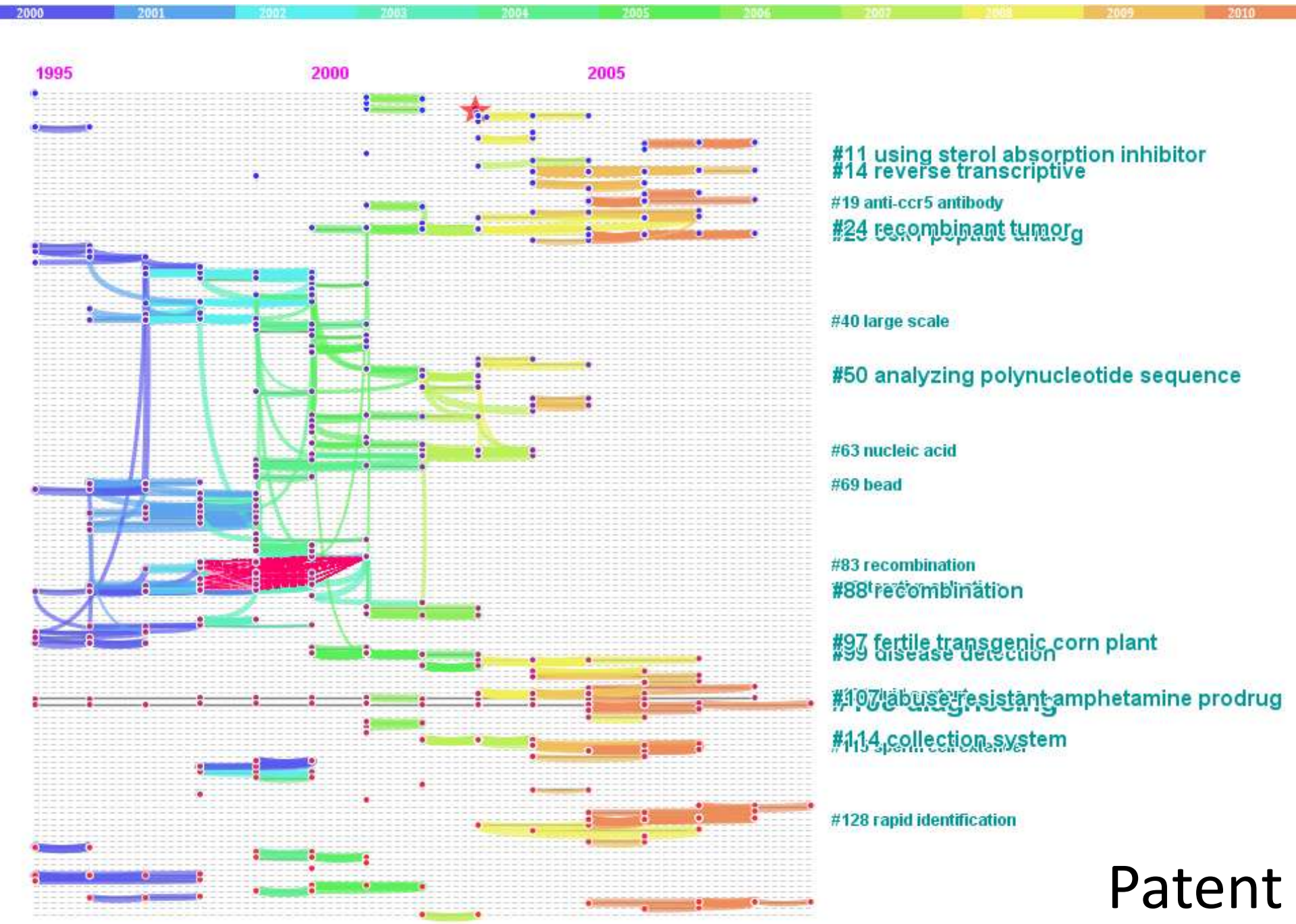
Go

Patent number: 6537746
Filing date: Dec 4, 1998
Issue date: Mar 25, 2003
Application number: 09/205,448

The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.

Inventors: Frances Arnold, Zhixin Shao, Alexander Volkov
Original Assignee: Maxygen, Inc.
Primary Examiner: W. Gary Jones
Secondary Examiner: Janell E. Taylor
Attorney: Townsend and Townsend and Crew LLP
Current U.S. Classification: 435/6.12; 435/69.1; 435/91.1; 435/91.2; 435/320.1; 506/9; 536/23.7
International Classification: C12Q/168

View patent at USPTO
Search USPTO Assignment Database



Patent

US6537746

A Visual Analytic Approach to Portfolio Analysis

- Boundary spanning and bridge building as early signs and explanatory mechanisms of transformative potential
- The capability of tracing the long-term negative impact of retracted articles as an integral part of portfolio analysis
- More in-depth case studies to identify the scope of the structural variation theory
- Automatically generate representations of knowledge at predicate levels