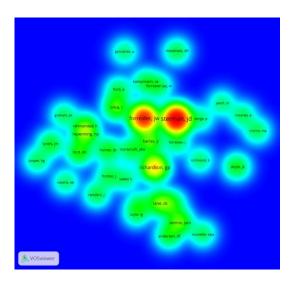
Análise Bibliométrica - Conceitos, Métodos e Softwares

Presenta	ation · April 2017	
DOI: 10.1314	0/RG.2.2.36548.01928	
CITATION		READS
1		2,553
1 author	:	
	Pedro Nascimento de Lima	
	Universidade do Vale do Rio dos Sinos	
	17 PUBLICATIONS 8 CITATIONS	
	SEE PROFILE	







ANÁLISE BIBLIOMÉTRICA

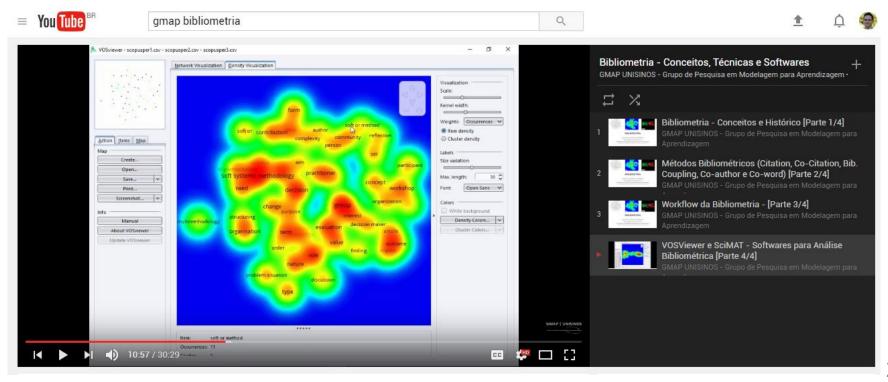
Conceitos, Técnicas e Softwares
Pedro Nascimento de Lima
Disciplina de Métodos Qualitativos de Pesquisa
– Prof. Daniel Pacheco Lacerda

Programa de Pós Graduação em Engenharia de Produção e Sistemas - UNISINOS



Esta apresentação foi gravada

• Link: https://www.youtube.com/watch?v=E4Ej746hrVM&list=PLgXXv1Y63Bsw-91pX2c5oMCOq0VoMnW6R&index=1



JESUÍTAS BRASIL

Agenda

- Parte 1: Histórico, Conceitos, Aplicações (~ 20 min);
- Parte 2: Workflow e Métodos (~ 60 min);
 - Workflow da Bibliometria e Métricas;
 - Métodos: Citation, Co-citation, Bibliographic Coupling, Co-Author e Co-Word;
- Parte 3: Softwares para Análise Bibliométrica (~ 40 min):
 - SciMAT & VOSviewer;

Bibliometria

PARTE 1 – HISTÓRICO, CONCEITOS E APLICAÇÕES





Bibliometria

- O que é*:
 "Bibliometria é a análise estatística de livros, artigos e outras publicações" (Glossário da OCED de Termos Estatísticos);
- Termos Similares: Cientometria (Scientometrics), Infometria (Informetrics).

^{*}Não existe uma definição apenas destes termos, e uma discussão aprofundada destes conceitos é realizada por Hood e Wilson (2001).

Bibliometria vs Cientometria

Bibliometrics vs Scientometrics

- Bibliometria: Foca principalmente em medir o output tangível dos cientistas (artigos, livros, patentes);
- Cientometria: Mais ampla do que a bibliometria, ocupa-se da análise quantitativa do desenvolvimento da pesquisa científica;
- Ambas são quantitativas.
- Mais sobre: Scientometrics Journal:
 - http://link.springer.com/journal/11192

Bibliometria - Origem

- Price (1965) Networks of Scientific Papers, Science, 149 (1965) 510-515;
- Usou estatística para estudar a natureza do desenvolvimento da ciência, como uma rede, usando a citação para relacionar dois papers.

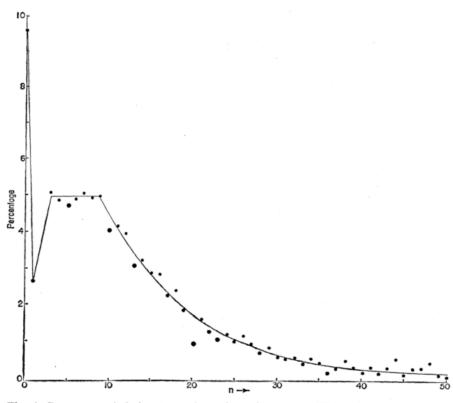


Fig. 1. Percentages (relative to total number of papers published in 1961) of papers published in 1961 which contain various numbers (n) of bibliographic references. The data, which represent a large sample, are from Garfield's 1961 *Index* (2).





Statistical Bibliography -> *Bibliometrics*Origem do Termo

 Pritchard (1969): Statistical Bibliography é um termo inadequado. A aplicação de métodos matemáticos e estatísticos à livros e outros meios de comunicação deve ser chamado de Bibliometria.





Bibliometria

Aplicações

- Bibliometria para medir "Performance Acadêmica";
- Bibliometria como campo de pesquisa em si;
- Bibliometria para entender a estrutura e o padrão de desenvolvimento da ciência;



Bibliometria – Medição de "Performance"

- Bibliometric / Citation Metrics: Avaliação de "Performance Acadêmica":
 - "Produtividade": Número de Papers publicados;
 - "Reconhecimento/Influência":
 - Número de Citações, em primeiro e segundo nível;
 - H-Index = 10 significa que de um corpo de trabalhos, pelo menos 10 trabalhos receberam 10 citações;
 - "Eficiência": Número médio de citações por artigo;
 Diversas outras métricas em Thomsom Reuters (2008).

Bibliometria – Campo de Conhecimento

Wood e Wilson (2001)

Table 2

Top twenty most productive journals in the literature of the three metric fields based on DIALOG's ranking *with* collapsing of journals due to variant forms of representation

Rank	No. records	Journal name (JN)	Language
1	1197	SCIENTOMETRICS	English
2	319	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE	English
3	285	NAUCHNO- TEKHNICHESKAYA INFORMATSIYA SERIES 1 & 2*	Russian & (Eng. Transl)
4	128	INFORMATION PROCESSING & MANAGEMENT	English
5	127	JOURNAL OF INFORMATION SCIENCE	English
6	109	JOURNAL OF DOCUMENTATION	English
7	95	REVISTA ESPANOLA DE DOCUMENTACION	



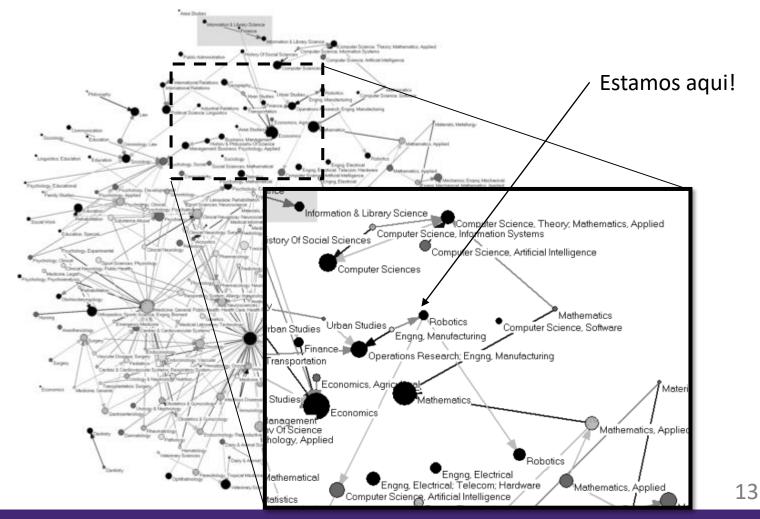
Bibliometria

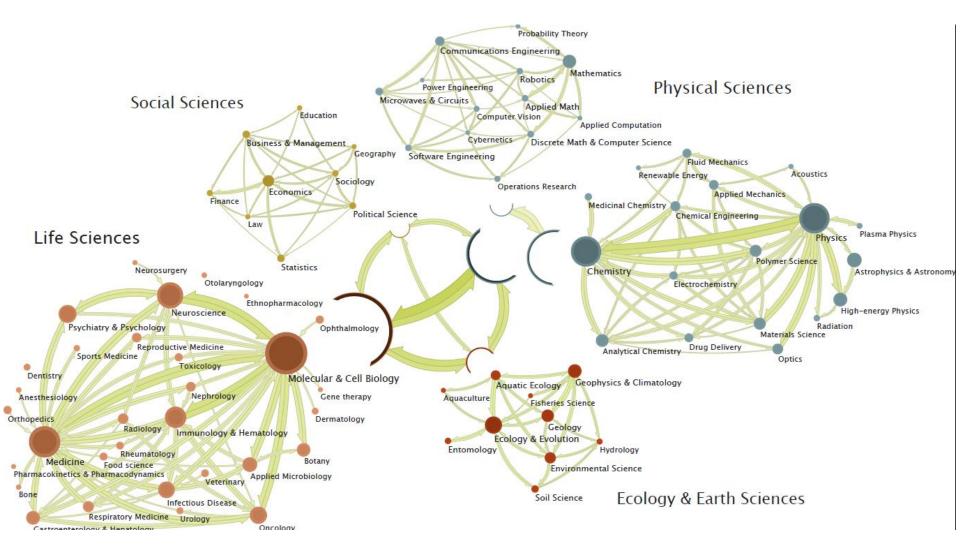
MÉTODOS PARA MAPEAR A CIÊNCIA



Entendendo a Estrutura

"Mapping the backbone of Science" – Boyack, Klavans, Borner (2005)





"Mapping Science" – http://www.eigenfactor.org/map/maps.php

www.gmap.unisinos.br

Métodos Bibliométricos para Mapear a Ciência

- O que eles fazem?
 - Relacionam trabalhos, autores, instituições, países, palavras chave de acordo com alguma métrica;
 - Mapeiam estas "unidades de análise" de acordo com a intensidade desta métrica;
 - Inferem insights a partir destas informações.
- Bibliometria é só isso?
 - Não. Pela sua definição ampla, no limite, qualquer conta que use dados bibliográficos é bibliometria.



Duas Abordagens Selecionadas

- Zupic e Ater (Essencial para usar o VosViewer)
 - Focam no que é essencial: Os Métodos Bibliométricos, e menos nas métricas e algoritmos.
- Cobo et al: Propõe abordagem para mapear o progresso do conhecimento em um campo. (Essencial para usar o SciMAT):
 - Mostram a bibliometria como um processo;
 - Criam o SciMAT para viabilizar o processo.

Passos para Mapear a Ciência com Métodos Bibliométricos

1. Projeto da Pesquisa

- Qual é a questão de pesquisa (base conceitual, front de pesquisa ou rede da pesquisa);
- Escolher o método apropriado

2. Compilar dados bibliométricos

- Selecionar / Construir base;
- Filtrar e exportar os dados bibliográficos.

3. Análise

- Escolher software de Análise Apropriado;
- Limpar os dados;
- Se necessário, construir matriz de similaridade para o software estatístico;
- Identificar subgrupos com o método escolhido (MDS, PCA, cluster analysis, network analysis);

4. Visualização

- Escolher o método de visualização adequado (MDS, Network Analysis);
- Selecionar o software adequado para visualização (UCINET< Pajek, Gephi, R iGraph, Citespace II, ...)

5. Interpretação

(Zupic e Ater 2014) 17





Métodos Bibliométricos abordados

Técnica	Por	Relação
Bibliographic Coupling	Autor, Trabalho ou Periódico	Referências comuns a um mesmo autor
Co-Author	Autor, País ou Instituição	Co-autoria
Co-Citation	Autor, Trabalho ou Periódico	Citações diferentes em um mesmo trabalho
Co-Word	Palavra chave ou palavra extraída do abstract	Ocorrência conjunta de dois termos

(Cobo et al 2011)

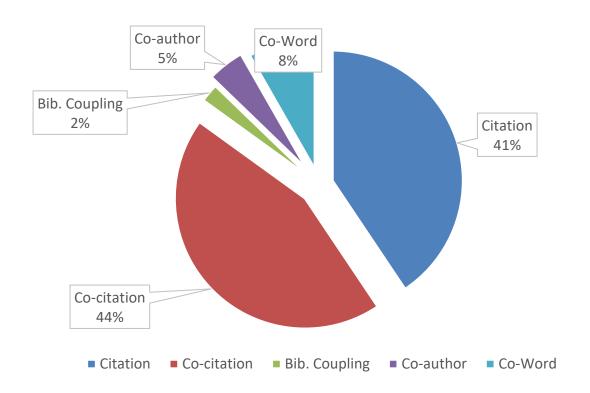
Métodos Bibliométricos abordados a seguir

- Principais Métodos Bibliométricos usados em Management (Zupic e Ater 2014):
 - Citation Analysis;
 - Cocitation-Analysis;
 - Bibliographic Coupling;
 - Co-author Analysis;
 - Co-Word.



Quais métodos são mais usados?

81 Trabalhos Analisados por Zupic e Ater (2014) em Management



(Zupic e Ater 2014) 20



Perguntas que pode ajudar a Responder

- Que autores mais influenciaram a pesquisa em um periódico?
- Que periódicos e disciplinas tem mais impacto em uma vertente de pesquisa?
- Quem são os experts em uma área de pesquisa?
- O que ler de uma determinada área?

(Zupic e Ater 2014)





- Usa a citação como uma medida de influência;
- Pressuposto: Autores, Trabalhos e Periódicos mais citados são mais influentes;





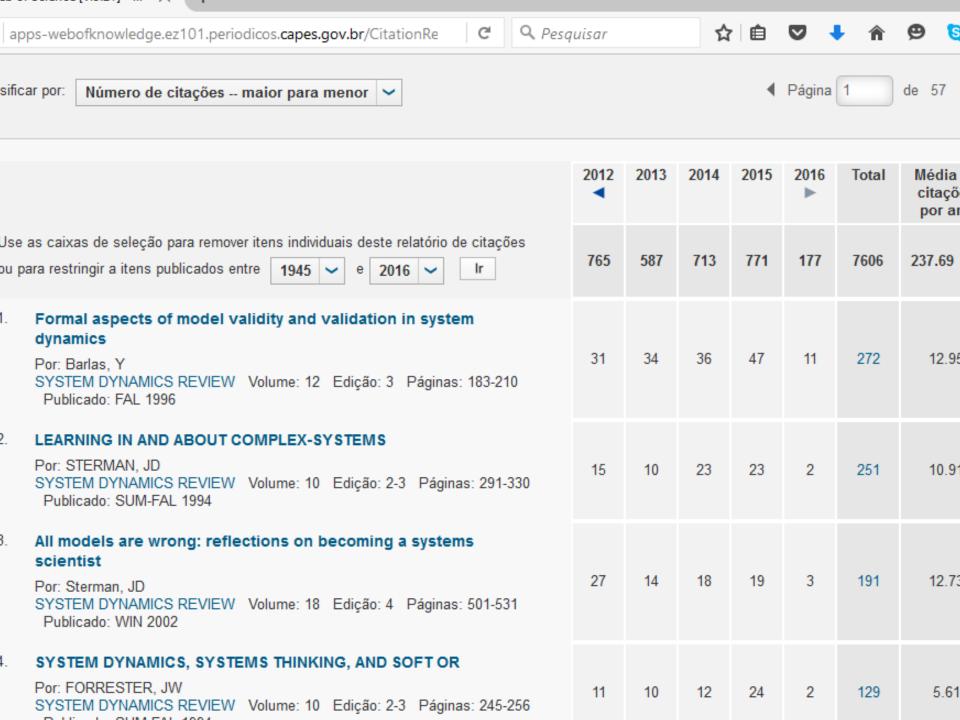
- Usa a citação como uma medida de influência;
- Pressuposto: Autores, Trabalhos e Periódicos mais citados são mais influentes;





Quais são os trabalhos influentes no System Dynamics Review?





Onde isso é pesquisado?

Que Organizações estão envolvidas nestes trabalhos?



wcs-webofknowled	ge.ez101.periodicos. capes.gov.br /RA/analy	yze. C	Q Pesqui	sar	☆	Ê		+	^	9	8	≡
lse as caixas de seleção	abaixo para visualizar os registros. É possív	el optar por vi	sualizar os re	gistros selecio	nados ou	exclui	-los (e	visual	izar os	outro	s).	
→ Visualizar registros							5	Salvar dados de análise no				
× Excluir registros	Campo: Organizações - Aprimorada		Contagem do registro % de 568		Gráfico de barras			Colunas de dados exiTodas as linhas de da				
	MASSACHUSETTS INSTITUTE OF TECHNOLOGY MIT	7	2	12.676 %	-							
	STATE UNIVERSITY OF NEW YORK SUNY SYSTEM	3	8	6.690 %	•							
	STATE UNIVERSITY OF NEW YORK SUNY ALBANY	3	7	6.514 %	•							
	UNIVERSITY OF LONDON	1	7	2.993 %	1							
	WORCESTER POLYTECHNIC INSTITUTE	1	7	2.993 %	1							
	UNIVERSITY OF BERGEN	1	5	2.641 %	1							
	UNIVERSITY OF MANNHEIM	1	5	2.641 %	1							
	ASIAN INSTITUTE OF TECHNOLOGY	1	3	2.289 %	1							
	RADBOUD UNIVERSITY NIJMEGEN	1	3	2.289 %	1							
	LONDON BUSINESS SCHOOL	1	0	1.761 %	1							
→ Visualizar registros								5	Salvar	dados	de análi	se no
× Excluir registros	Campo: Organizações - Aprimorada	Contagem	do registro	% de 568	Gráfico	o de b	arras	\sim			dados	
	(114 valor(es) de Organizações - Aprimor (83 registros(14.613%) não contêm dados r											
2016 THOMSON REUTERS	TERMOS DE USO POLÍTICA DE PRIVA	CIDADE (COMENTÁRIOS									

And intermediated by an explanation of the district of the dis

Perguntas que pode ajudar a Responder

- Quem são as organizações centrais e periféricos neste campo de pesquisa?
- Qual é o grupo de autores citado sistematicamente por um grupo determinado de trabalhos?
- Que trabalhos são referenciados conjuntamente?

(Zupic e Ater 2014)





- Usa a citação a dois trabalhos como uma medida de similaridade entre trabalhos citados, autores ou periódicos;
- Pressuposto: Quanto mais dois trabalhos são citados juntos, mais seu conteúdo está relacionado;
- Pode ser feito por trabalho, autor ou periódicos.
- *Tri-Citation*: Usa um autor focal e analisa os pares que citaram-se mutuamente com o mesmo.

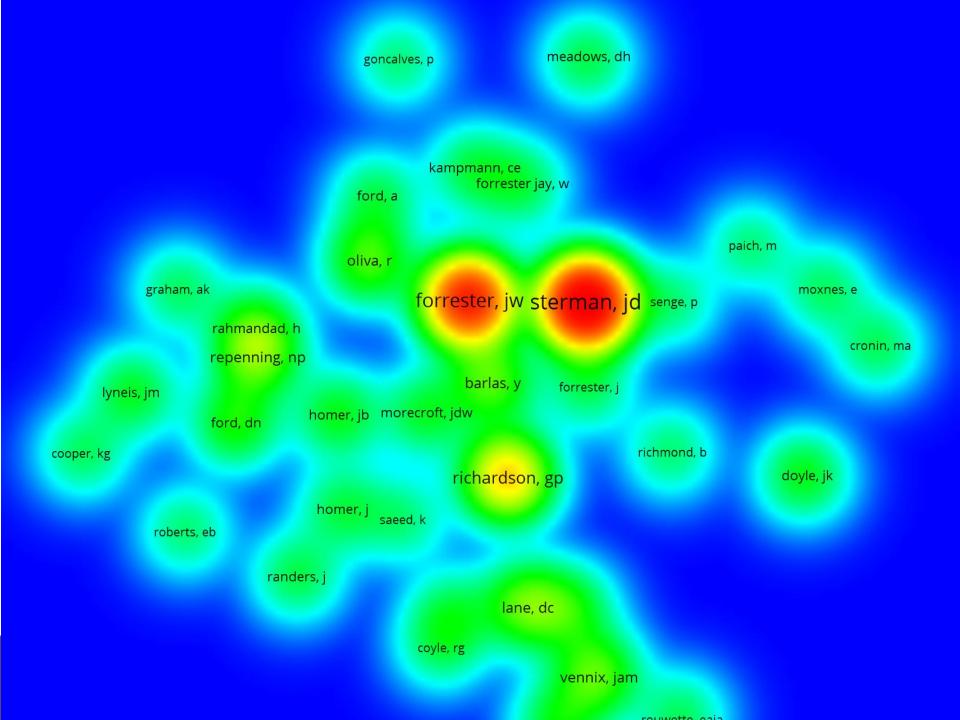




Co-Citation

Quais são os Autores mais Influentes na área da Dinâmica de Sistemas, e quais são citados *juntos*?





Co-Citation

Quais são os trabalhos mais influentes em dinâmica de sistemas? Que trabalhos são citados junto?



repenning np, 2001, calif mana

forrester jay w., 1980, tims s

n m, 1993, manage sci, v39

forrester j. w., 1961, ind dyn

erman jd, 1989, organ behav

sterman j. d., 2000, business

sterman jd, 2002, syst dynam r

vennix j. a. m., 1996, group m

eeney lb, <mark>2000, s</mark>yst dynam r

sterman jd, 1994, syst dynam r

Bibliographic Coupling

Perguntas que pode ajudar a Responder

- Quem são as organizações centrais e periféricos em um grupo de pesquisa emergente?
- Que autores produziram sínteses que aproximaram um corpo da literatura?
- Como a estrutura de uma vertente de pesquisa reflete a diversidade de abordagens teóricas.

(Zupic e Ater 2014)





Bibliographic Coupling

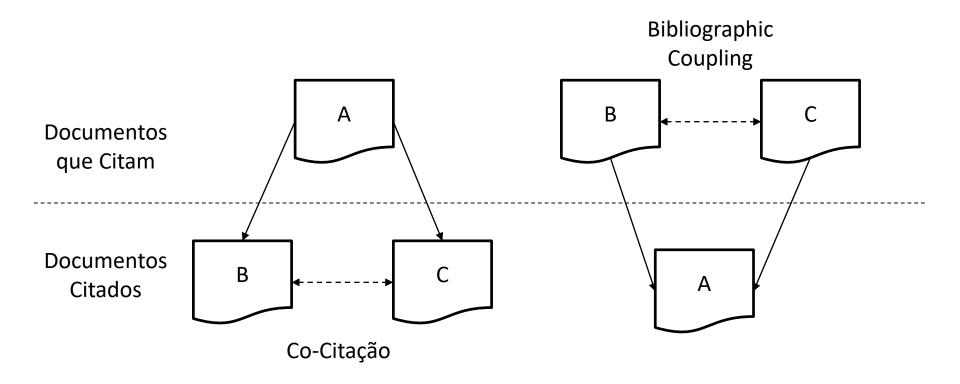
- Usa o número de referências compartilhadas por dois trabalhos como uma medida de similaridade entre eles.
- Pressuposto: Quanto mais dois trabalhos citam trabalhos parecidos, mais seu conteúdo está relacionado;
- Apenas faz sentido se os artigos pesquisados forem de um período temporal específico (Não faz sentido comparar as referências de um trabalho de 1900 com as referências de um trabalho atual);

(Zupic e Ater 2014) 35





Co-Citation vs Bibliographic Coupling



(Zupic e Ater 2014) 36

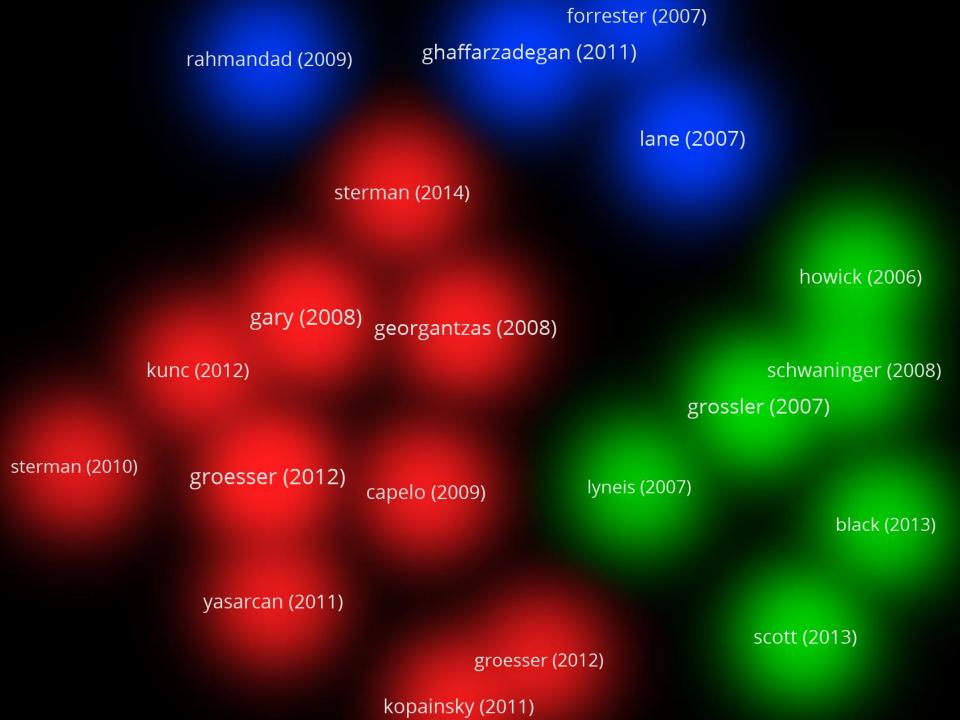




Bibliographic Coupling

Quais são os trabalhos e últimos fronts de pesquisa em dinâmica de sistemas?





Bibliographic Coupling

E em Design Science?



E JESUÍTAS BRASIL



```
helfert m.
                     eurich m.
                                 meyer m. ghaisas s.
                                                               kristiansen e.
                                hammell r.j.
                       king g.
                                                                                                                       koskela
                                   palmqvist h.
                                                 dresch a. athanasoulis m.
                                                                                                               canters k.j.
                                                                                      avenier m.-j.
                                                                        berthod o.
                                      ven k.
                                                                                                       hendriks I.
                                                                                           hyötyläinen r. cash p.
            o'flaherty b.
                                                                                     stam c.d.
                            mettler t. júnior j.a.v.a.
                                                                                              tzortzopoulos p.
                                              ostrowski ł. curley m. van haaster k.j.m.
                                                                                   march s.t.
                         schwabe g. alturki a. venable j.
   bernstein a.
                 thornton c.
                                          akoka j. karow m.
                                                              hevner a.r.
                               riedl r.
                          wang h. winter r. verelst j. marshall p.
                baptista a.a.
                                                 tuunanen t. brady d.a.
         gama n.
                                    proper h.a.
                                                               khazanchi d.
                             shrestha a. toleman m.
                                                                                          arnott d.
                                                              birkmeier d.g.
                 da silva m.m.
                                breitner m.h. becker j.
                                      olbrich s.
                     lahrmann g.
                                                             buckl s.
                             recker j. krcmar h. teuteberg f.
                        wortmann f.
    frank u.
                                                             barclay c.
                                shi y. goes p.b. kerr d.
                                    alter s. müller o.
                       lau r.y.k. cavaco a. brenner w.
                                     long c. dehling t.
                      xu k.
```

Co-Citation vs Bibliographic Coupling

- Bibliographic Coupling: Mais adequado para mapear o front de pesquisa atual;
- Co-citation: Mais adequado para mapear artigos seminais e antigos.

(Zupic e Ater 2014) 41



Limitações destas Análises

- Apenas com a Citação, não é possível saber porque um trabalho foi citado (ex.: Para confirmar ou refutar uma teoria);
- Inevitavelmente irá carregar todo o viés possível embutido nas citações.

Co-Author Analysis

Perguntas que pode ajudar a Responder

- Que pesquisadores trabalham junto?
- Como é a estrutura Social de uma área de pesquisa?
- Que instituições ou países colaboram neste campo de pesquisa?

(Zupic e Ater 2014)





Co-Author Analysis

- Examina o aspecto Social da pesquisa, ao invés do aspecto técnico;
- Usa uma medida de colaboração;
- Pode ser usado para avaliar a colaboração em nível institucional e nacional.





Co-Author Analysis

Quem publica junto em dinâmica de sistemas?





ghaffarzadegan, n

richardson, gp

andersen, d

Co-Word Analysis

Perguntas que pode ajudar a Responder

- Que palavras-chave vem sendo mais usadas em cada período determinado de tempo?
- Que palavras são usadas em conjunto?
- Como o interesse de pesquisa mudou?





Co-Word Analysis

- Técnica da Análise de Conteúdo;
- Pressuposto: Se as palavras se repetem em diversos documentos, os conceitos relacionados à estas palavras está relacionado;
- Pode ser aplicado ao título, abstract, palavraschave ou ao texto completo;
- Output: Rede de Temas;

(Zupic e Ater 2014) 48







Co-Word Analysis - Limitações

- Com base em Palavras-Chave:
 - As bases não tem necessariamente as palavraschave corretamente definidas;
 - As palavras-chave podem não representar adequadamente o tema;
- Com base no texto ou Abstract:
 - Carrega ruído para a análise.

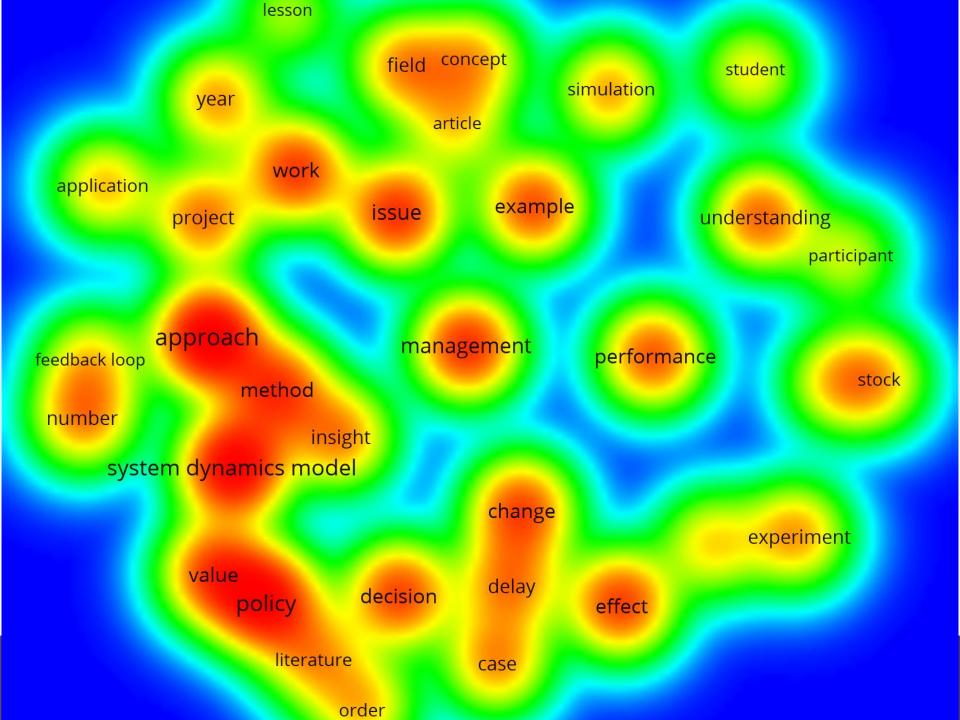
(Zupic e Ater 2014) 49

Co-Word Analysis

Que palavras chave são mais usadas no System Dynamics Review?







Co-Word Analysis

E no EJOR?

(Zupic e Ater 2014) 52





resilience

supply chain
supplier
firm

competitive advantage

organization

performance relationship collaboration

financial performance

profitability

team autonomy

Co-Word Analysis

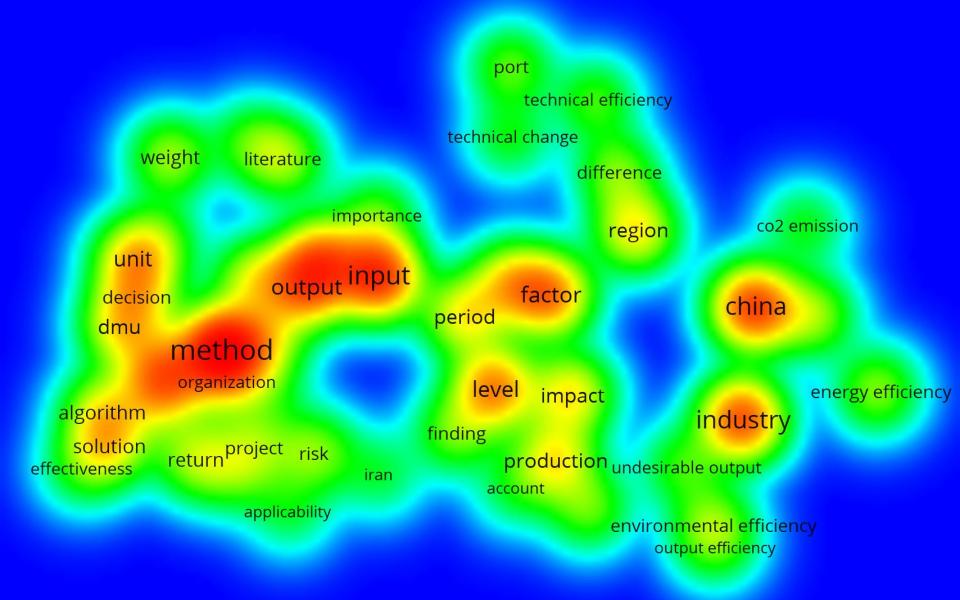
Que outras palavras estão sendo usadas com o DEA (Data Envelopment Analysis em 2016)?

(Zupic e Ater 2014) 54







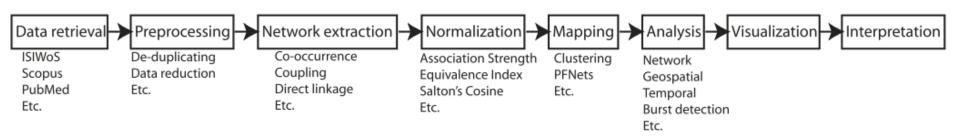


Resumo – Prós e Contras

Método (Unidades de Análise)	Pontos Fortes	Pontos Fracos
Citation (<i>Trabalho, Autor,</i> <i>Periódico</i>)	Acha facilmente trabalhos importantes de uma área	Novas publicações tem menos chance de serem consideradas importantes.
Co-Citation (<i>Trabalho, Autor,</i> <i>Periódico</i>)	Conectar trabalhos, autores e periódicos com co-citação é uma medida confiável.	Não é ótimo para mapear "research fronts", porque mapeia artigos <i>citados</i> .
Bibliographic Coupling (<i>Trabalho, Autor, Periódico</i>)	Pode ser usado para achar áreas de pesquisa novas, e sub-áreas menores.	Só pode ser usado em um timeframe reduzido (ex.: até 5 anos). É difícil saber se os trabalhos encontrados são importantes.
Co-author (Autor)	Mostra evidência de colaboração e a estrutura social do campo.	Nem sempre colaboração é apresentada com a coautoria.
Co-word G (Palavra)	Pode usar o <i>conteúdo</i> do artigo para análise.	Palavras aparecem em diferentes formas e com diferentes significados.

Workflow of Science Mapping

Uma visão mais Aberta



(Cobo et al. 2012) 57



Sobrevoando o processo

Nada é de graça

- Busca de Dados:
 - Web of Science;
 - Scopus;
 - Observar critérios para busca conforme uma Revisão Sistemática da Literatura;
 - Output: Lista de Trabalhos com citações.
- Pré-Processamento:
 - Eliminar Autores, Journals, etc com nomes duplicados / errados;
 - Redução de Dados: Escolher os itens mais relevantes de acordo com algum critério (Ex.: Número mínimo de citações).

(Cobo et al. 2011, 2012) 58





Sobrevoando o processo

Nada é de graça

- Network Extraction:
 - Escolha da Unidade de Análise: Autor, Periódico, Trabalho, etc.;
 - Escolha da Relação entre os nós: Co-ocorrência, coupling, ou referência;
- Normalização Calculando a Intensidade entre os nós da rede:
 - Salton Cosine;
 - Jaccard Index;
 - Equivalence Index;
 - Association strength.



Índice de Similaridade

• Índice de Similaridade entre dois itens:

$$s_{ij} = \frac{c_{ij}}{w_i w_j}$$

 s_{ij} : Similaridade entre o item i e j

c_{ij} : Coocorrências de i e j

w_i : Número de ocorrências i

w_i : Número de ocorrências j

Índice de Equivalência / Associação

• Índice de Similaridade entre dois itens:

$$e_{ij} = \frac{c_{ij}^2}{c_i c_j}$$

 e_{ij} : Associação entre o item i e j

c_{ij} : Coocorrências de i e j

 c_i : Número de ocorrências i

c_i: Número de ocorrências j

Sobrevoando o processo

Nada é de graça

- Construindo o Mapa
 - Existem diversos métodos para construir um mapa a partir de uma matriz:
 - Multidimensional scaling;
 - Algoritmos de Clusterização;
 - Algoritmo VOS (pelos criadores do VOSViewer);
- Métodos de Análise
 - Network Analysis;
 - Longitudinal Analysis;
 - Geospatial Analysis;
 - Performance Analysis;

(Cobo et al. 2012) 62

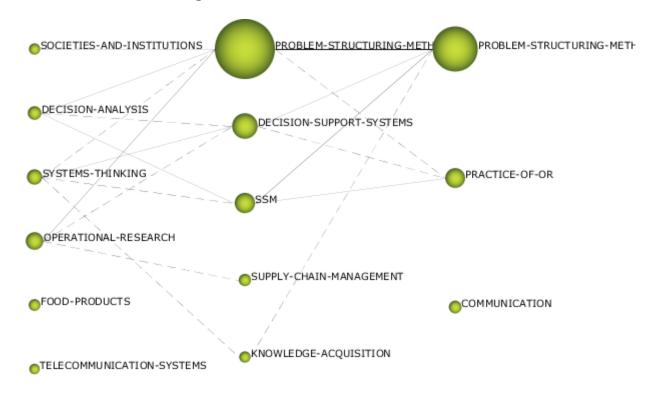




Longitudinal Analysis

Observando o Desenvolvimento dos Termos

Mapa de Evolução



Longitudinal Analysis

Observando o Desenvolvimento dos Termos

- Mapa de Evolução:
 - Relação entre os clusters U e V é calculada pelo Inclusion Index;
 - Tamanho do tema é definido por um critério de performance.

$$Inclusion\ Index = \frac{\# (U \cap V)}{\min(\# U, \# V)}$$

(Cobo et al. 2011) 64



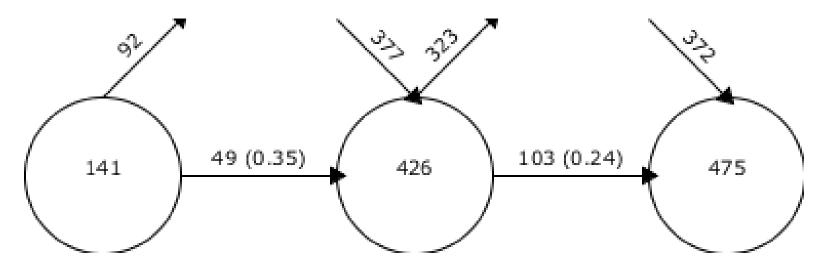




Longitudinal Analysis

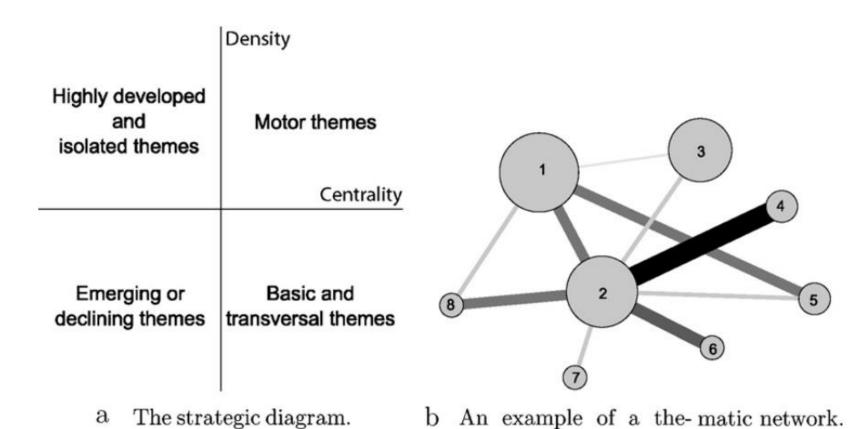
Observando o Desenvolvimento dos Termos

 Mapa de Sobreposição (Índice de Estabilidade)



(Cobo et al. 2012) 65



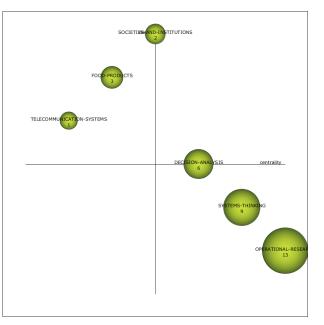


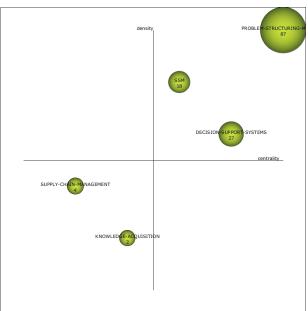
(Cobo et al. 2011) 66

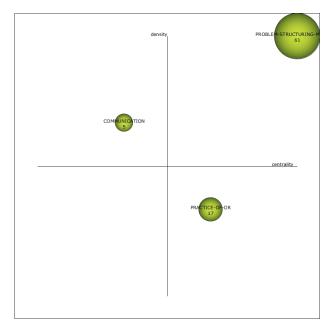




Diagrama Estratégico











Rede Temática de um Cluster

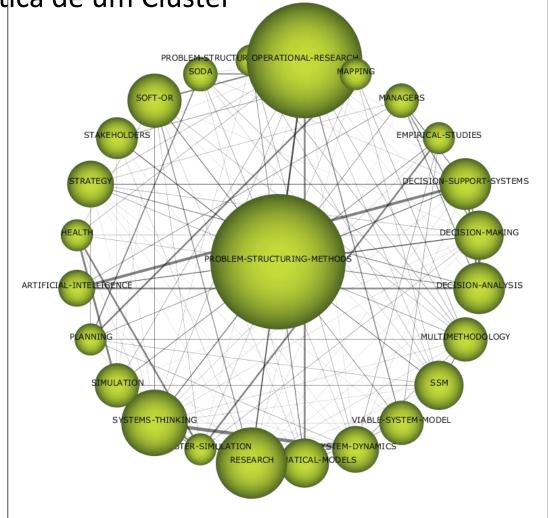




Diagrama Estratégico

- Callons's Centrality & Density:
 - Centralidade: Grau de Interação com os outros clusters;
 - Densidade: Força / Coesão interna do Cluster.

$$c = 10 * \sum e_{kh}$$
 $d = 100 \left(\frac{\sum e_{ij}}{w}\right)$

 e_{kh} : Índice de equivalência da palavra k do cluster com as palavras h do outro tema

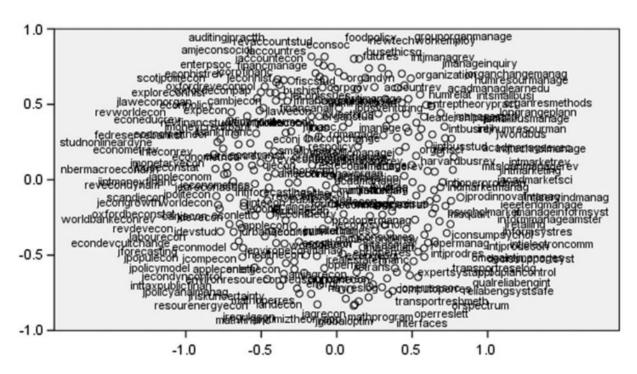
 e_{ij} : Índice de equivalência das palavras i e j do Cluster

w: Número de palavras do Cluster



Ok, Como Fazer?

Mapa com o SPSS:



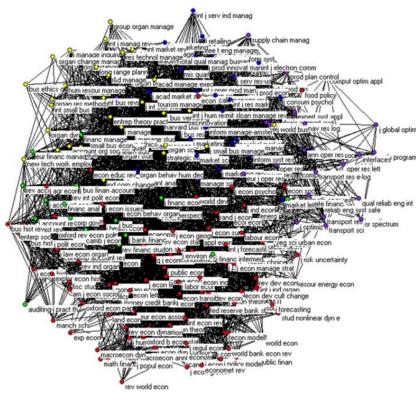
(van Eck e Waltman 2010) 70





Ok, Como Fazer?

Mapa com o Pajek:



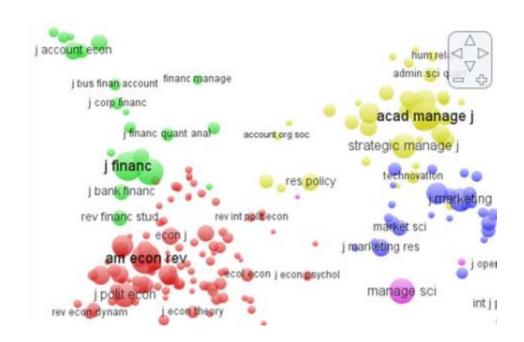
ıvan Eck e Waltman 2010) 71





Ok, Como Fazer?

Mapa com o vosViewer:



(van Eck e Waltman 2010) 72





Softwares para Bibliometria

O que não falta é opção

- Softwares Analisados (em 2011):
 - Bibexcel;
 - CiteSpace;
 - ColPalRed
 - IN-SPIRE;
 - Leydesdorff's Software;
 - Network Workbench Tool;
 - Science of Science Tool;
 - Vantage Point;
 - VOSViewer.
 - E SciMAT;





(Cobo et al. 2011)

Softwares que veremos aqui

- VOSViewer: Visualizações boas o suficiente em pouco tempo;
- SciMAT: Flexibilidade para quem quer se aprofundar.
- Cobo et al 2011: Discussão aprofundada de Prós e Contras de cada um dos 9 Softwares. Proposição de Método alinhado ao SciMAT;
- Cobo et al 2012: Propõe o SciMAT;

(Cobo et al. 2011)



E se eu quiser estudar mais?

- Cobo et. al (2011): Compararam 9 softwares diferentes para a realização de Science Mapping;
- Cobo et. al (2012): Criaram o Software SciMAT, citado por 59 estudos até hoje;
- **Zupic e Ater (2014)**: Sintetiza conhecimento sobre *como* realizar um estudo bibliométrico, as técnicas mais utilizadas, etc.

Referências

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for Information Science and Technology, 62(7), 1382–1402.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. Journal of the American Society for Information Science and Technology, 63(8), 1609–1630. http://doi.org/10.1002/asi.22688

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314. http://doi.org/10.1023/A:1017919924342

Price, D. J. de S. (1965). Networks of Scientific Papers, 149(3683), 510–515. http://doi.org/10.1126/science.149.3683.510

Pritchard, A. (1969). Statistical bibliography or bibliometrics. Journal of Documentation, 25, 348.

Thomsom Reuters. (2008). Whitepaper Using Bibliometrics: A guide to evaluating research performance with citation data. Thomson Reuters, 12.

http://doi.org/10.1097/NCN.0b013e31819ec9ac

Zupic, I., & Ater, T. (2014). Bibliometric Methods in Management and Organization. Organizational Research Methods, 18(3), 429–472. http://doi.org/10.1177/1094428114562629