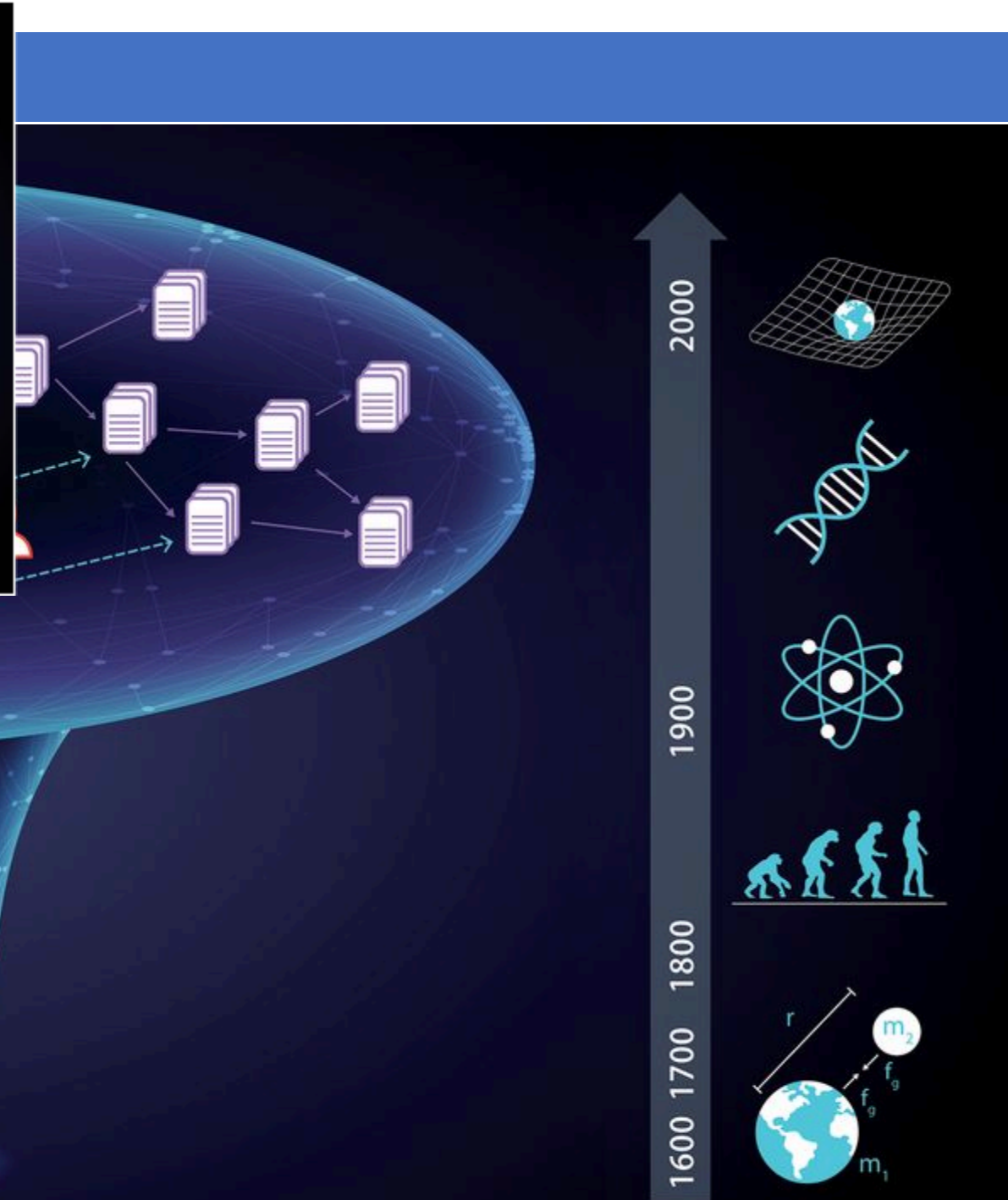
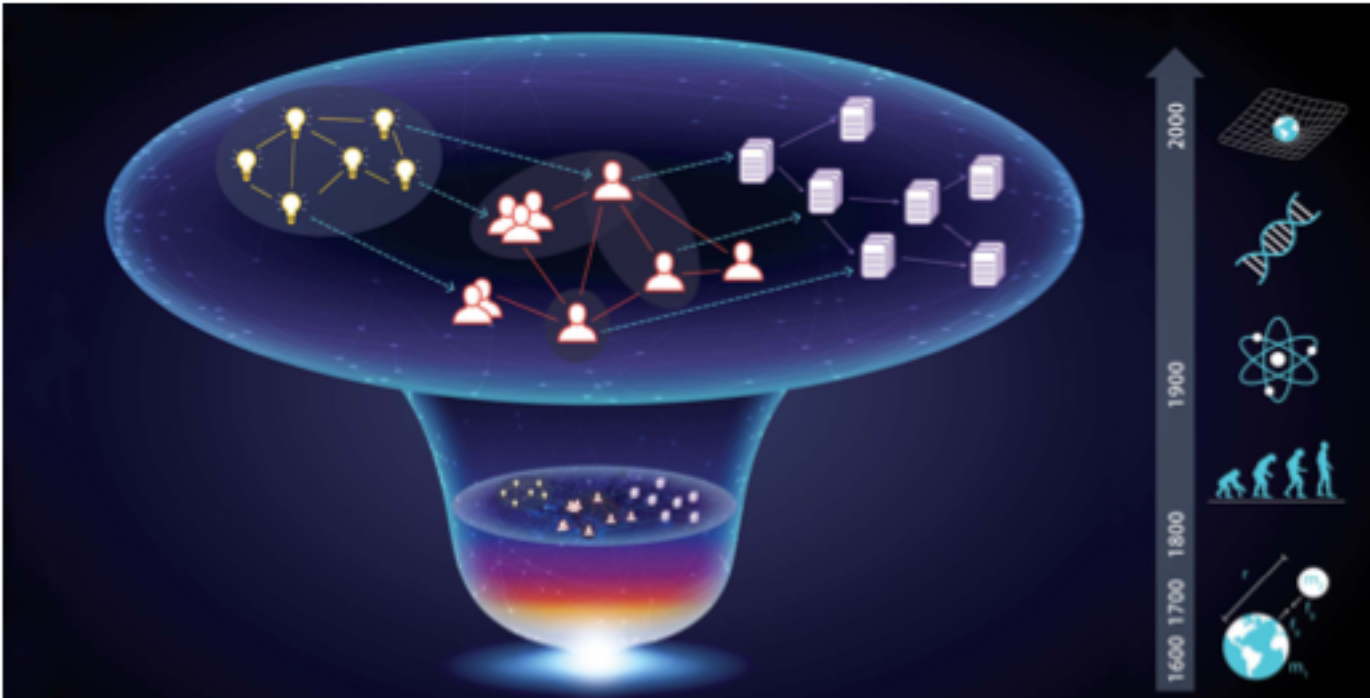


Science of Science (SciSci)

BY SANTO FORTUNATO, CARL T. BERGSTROM, KATY BÖRNER, JAMES A. EVANS, DIRK HELBING, STAŠA MILOJEVIĆ, ALEXANDER M. PETERSEN, FILIPPO RADICCHI, ROBERTA SINATRA, BRIAN UZZI, ALESSANDRO VESPIGNANI, LUDO WALTMAN, DASHUN WANG, ALBERT-LÁSZLÓ BARABÁSI
SCIENCE | 02 MAR 2018

- **Surgimento da SciSci**: a crescente disponibilidade de dados digitais sobre entradas e saídas acadêmicas - de **financiamento de pesquisa**, **produtividade** e **colaboração** a **citações de artigos** e **mobilidade de cientistas** - oferece oportunidades sem precedentes para explorar a estrutura e a evolução da ciência.
 - *Web of Science* (proprietária, o primeiro índice de citação histórico), *Scopus*, *Google Acadêmico* e outros.
- A SciSci coloca a própria ciência no microscópio oferecendo uma compreensão quantitativa das interações entre agentes científicos em diversas escalas geográficas e temporais.
- **Objetivo**: desenvolver ferramentas e políticas que tenham o potencial para acelerar o progresso científico.
- **Proposta da SciSci**: com uma compreensão mais profunda dos fatores que impulsionam a ciência de sucesso, podemos abordar de maneira mais eficaz problemas ambientais, sociais e tecnológicos.



A ciência pode ser vista como uma rede complexa auto-organizada e em evolução de ideias/projetos, acadêmicos e artigos. A SciSci procura por leis universais e específicas de domínio subjacentes à estrutura e dinâmica da ciência.

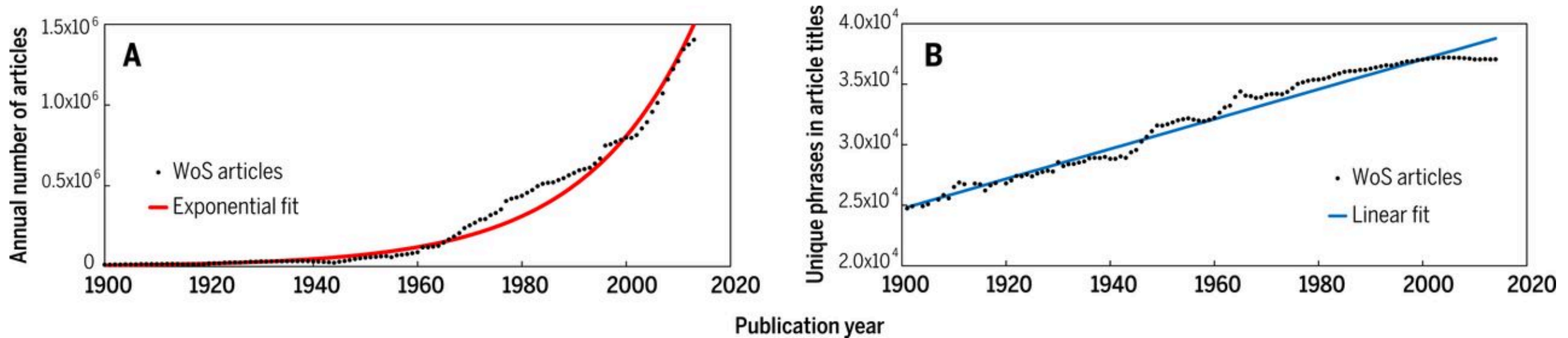
- Esta representação revelou padrões que caracterizam o surgimento de novos campos científicos através do estudo de **redes de colaboração** e o caminho de descobertas impactantes através do estudo de **redes de citações**.
- Modelos microscópicos traçaram a dinâmica da acumulação de citações, permitindo-nos prever o impacto futuro de artigos individuais.
- A SciSci revelou escolhas e *trade-offs* que os cientistas enfrentam à medida que avançam tanto em suas próprias carreiras quanto no horizonte científico.
 - Estudar tópicos relacionados aos seus **conhecimentos atuais** ou quebrar este padrão e se envolverem em **carreiras mais arriscadas**? (limitações no potencial de descobertas futuras)
 - Equipes **pequenas** ou **grades**? (influencia no impacto e duração)

- A SciSci oferece uma profunda **compreensão quantitativa** da estrutura relacional entre cientistas, instituições e ideias, porque facilita a identificação de mecanismos fundamentais responsáveis pela descoberta científica.
- Embora a SciSci busque leis e mecanismos universais de longa data que se aplicam a vários campos da ciência, um desafio fundamental daqui para frente é explicar diferenças inegáveis de cultura, hábitos e preferências entre diferentes campos e países.
 - Essa variação dificulta a compreensão de alguns insights de vários domínios e dificulta a implementação de políticas científicas associadas.
- As diferenças entre as questões, os dados e as habilidades específicas de cada disciplina sugerem que mais *insights* podem ser obtidos a partir de estudos SciSci específicos de domínio, que modelam e identificam oportunidades adaptadas às necessidades de campos de pesquisa individuais.

1. NETWORKS OF SCIENTISTS, INSTITUTIONS, AND IDEAS

- O **conhecimento científico** é constituído por conceitos e relações incorporados em artigos de pesquisa, livros, patentes, *software* e outros artefatos acadêmicos, organizados em disciplinas científicas e campos mais amplos.
- **Não temos como equiparar o crescimento da literatura científica ao crescimento de ideias científicas.**
- Crescimento exponencial no volume da **literatura científica**, uma tendência que continua com um período médio de duplicação de 15 anos.
 - Mudanças no mundo editorial, tanto tecnológico quanto econômico, levaram ao aumento da eficiência na produção de publicações.
 - Novas publicações na ciência tendem a se agrupar em áreas distintas do conhecimento.
- O território conceitual da ciência se expande linearmente com o tempo.

GROWTH OF SCIENCE



(A) Produção anual de artigos científicos indexados na base de dados do WoS. Podemos notar um crescimento exponencial no volume da literatura científica, uma tendência que continua com um período médio de duplicação de 15 anos.

(B) Crescimento de ideias cobertas por artigos indexados no WoS. Isso foi determinado pela contagem de frases extraídas de títulos e resumos para medir a extensão cognitiva da literatura científica.

Enquanto o número de publicações cresce exponencialmente, o número de ideias científicas se expande apenas linearmente.

1. NETWORKS OF SCIENTISTS, INSTITUTIONS, AND IDEAS

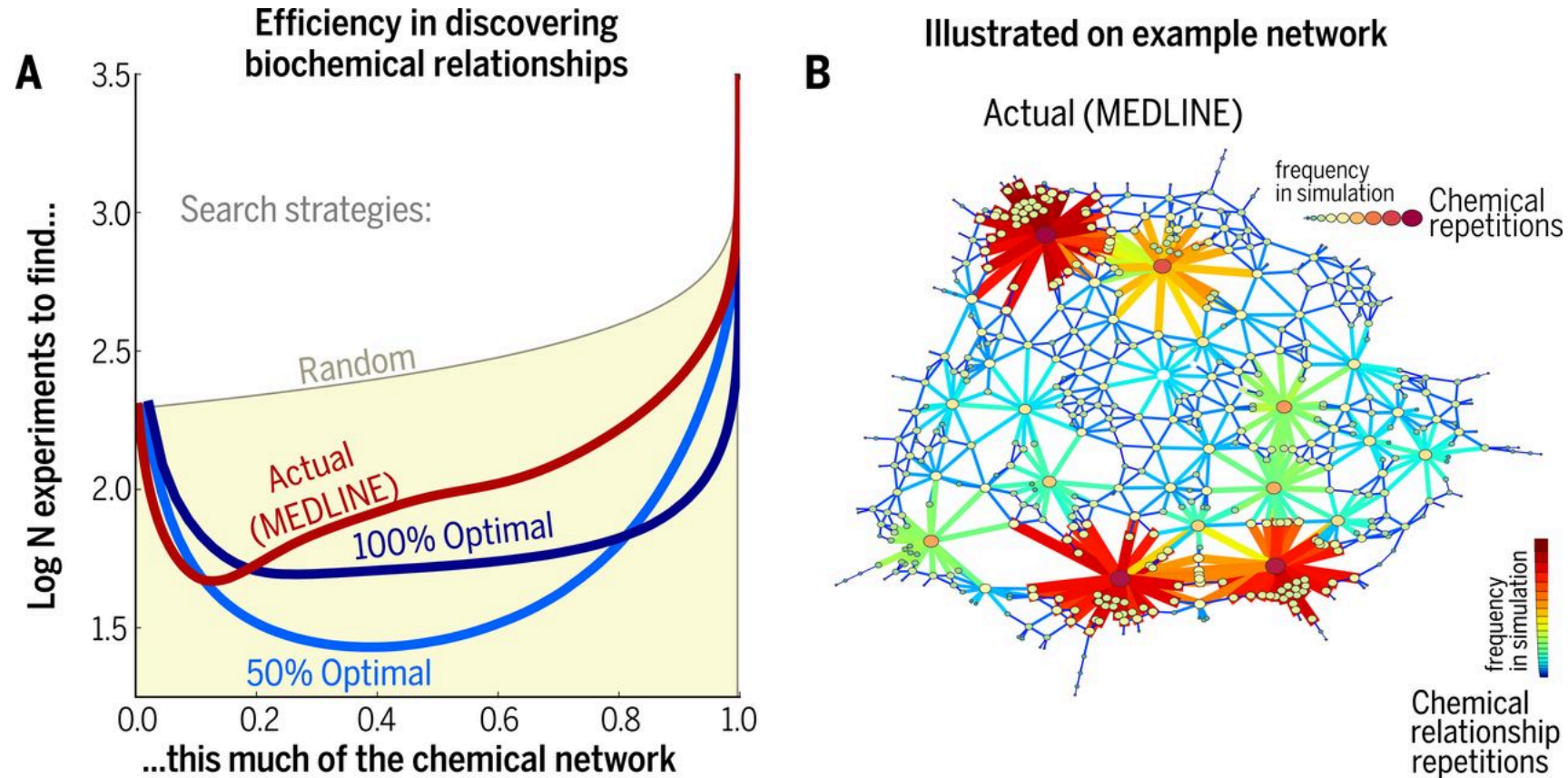
- Palavras e frases que ocorrem com frequência em títulos de artigos e resumos se propagam através de **redes de citações**, pontuadas por surtos correspondentes ao surgimento de novos paradigmas.
- Ao aplicar os métodos da ciência de redes às **redes de citações**, os pesquisadores são capazes de identificar comunidades definidas por subconjuntos de publicações que frequentemente se referem uns aos outros.
 - Grupos de autores que mantêm uma posição comum em relação a questões específicas ou trabalhando nos mesmos subtópicos especializados.
 - O crescimento da literatura reforça essas comunidades.
- À medida que novos artigos são publicados, são adicionados *links* entre cientistas e “coisas”, que são os nós da rede. A maioria dos novos links está entre as “coisas” a uma distância de apenas um ou dois passos, o que implica que, quando os cientistas escolhem novos tópicos, eles preferem “coisas” diretamente relacionadas à sua expertise atual ou a dos seus colaboradores. Essa densificação sugere que **a estrutura existente da ciência pode restringir o que será estudado no futuro**.

2. PROBLEM SELECTION

- **Como os cientistas decidem quais problemas de pesquisa trabalhar?**
 - Essas escolhas são moldadas por uma tensão contínua entre pesquisa conservadora e inovação arriscada.
- Cientistas que preferem uma pesquisa mais tradicional, muitas vezes parecem produtivos, publicando um fluxo constante de contribuições que avançam uma agenda de pesquisa focada.
 - Pode limitar a capacidade do pesquisador de sentir e aproveitar as oportunidades para estabelecer novas ideias que são necessárias para ampliar o conhecimento do campo.
- Um estudo de caso com foco em cientistas biomédicos, escolhendo novas substâncias químicas e relações químicas, mostra que à medida que os campos amadurecem, os pesquisadores tendem a se concentrar cada vez mais no conhecimento estabelecido.

CHOOSING EXPERIMENTS TO ACCELERATE COLLECTIVE DISCOVERY

As escolhas coletivas dos cientistas determinam mais amplamente a direção da descoberta científica.



- (A) A taxa de eficiência média para estratégias científicas globais para descobrir novas relações químicas publicáveis, estimadas a partir de todos os artigos indexados no MEDLINE publicados em 2010.
- (B) O processo de pesquisa real e estimada, ilustrado em uma rede hipotética de relações químicas, obteve a média de 500 execuções simuladas dessa estratégia.

2. PROBLEM SELECTION

- A **publicação inovadora** tende a resultar em um maior impacto do que uma **publicação conservadora**.
 - No entanto, as estratégias de inovação de alto risco são raras, porque a recompensa adicional não compensa o risco de falha na publicação.
 - Prêmios e reconhecimentos científicos funcionam como **incentivos**.
- A escolha de problemas de pesquisa pelos cientistas afeta principalmente suas carreiras individuais e as carreiras daqueles que dependem delas.
- Estratégias conservadoras servem bem as carreiras individuais, mas são menos eficazes para a ciência como um todo.
 - Hipóteses mais arriscadas podem ter sido testadas por gerações de cientistas, mas apenas aquelas bem sucedidas o suficiente para resultar em publicações são conhecidas por nós.
 - **Solução:** estimular as agências financiadoras a patrocinar proativamente projetos arriscados que testem hipóteses realmente inexploradas e que aceitem grupos de interesse especial.

3. NOVELTY

- Análises de publicações e patentes revelam consistentemente que combinações raras em descobertas científicas e invenções tendem a obter **maiores taxas de citação**.
- A **ciência de maior impacto** é fundamentalmente baseada em combinações convencionais de trabalho anterior, mas simultaneamente apresenta combinações incomuns. Artigos deste tipo são duas vezes mais propensos a receberem altas citações.
- **Em outras palavras, uma mistura equilibrada de elementos novos e estabelecidos é o caminho mais seguro para a aceitação bem-sucedida dos avanços científicos.**

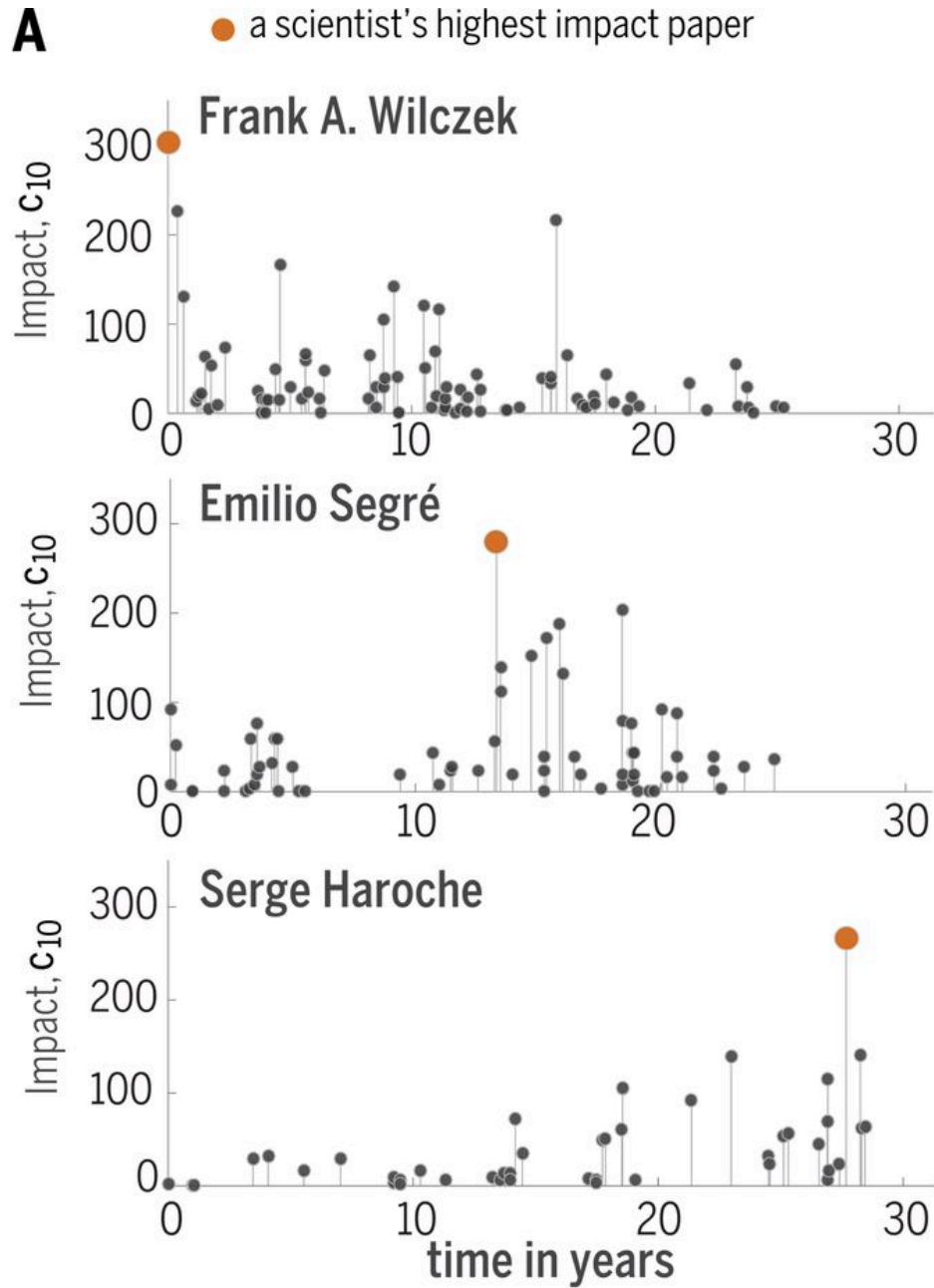
4. CAREER DYNAMICS

- **Carreiras acadêmicas individuais** se desdobram no contexto de um vasto mercado para produção e consumo de conhecimento.
 - Examinadas em termos não só de incentivos individuais e produtividade marginal, mas também de incentivos institucionais e competição.
- Esquemas de **financiamento tolerantes a falhas precoces**, que recompensam o sucesso a longo prazo, têm maior probabilidade de gerar publicações de alto impacto do que os subsídios sujeitos a ciclos de revisão curtos.
- A **desigualdade de gênero** na ciência continua predominante e problemática.
 - As mulheres têm menos publicações e colaboradores e menos financiamento, e são penalizadas nas decisões de contratação quando comparadas a homens igualmente qualificados.
 - Evidências experimentais mostram que os preconceitos contra as mulheres ocorrem nos estágios iniciais da carreira.

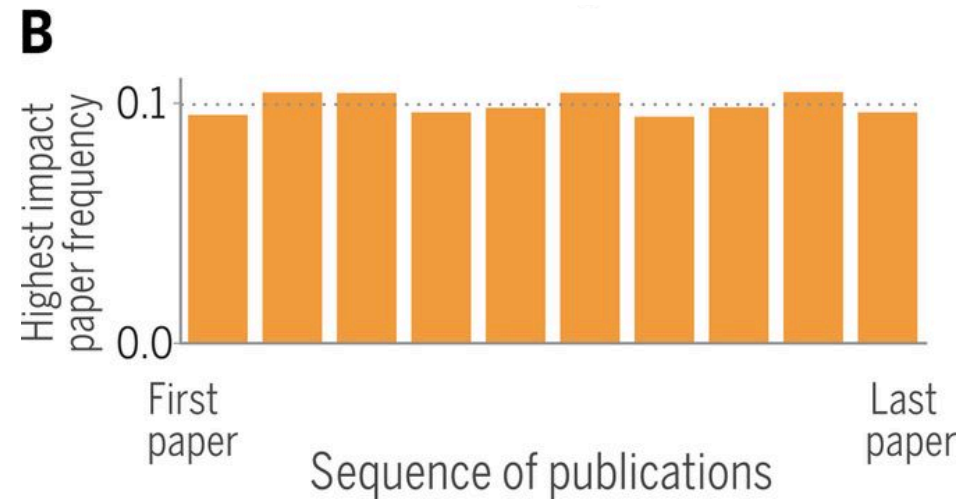
4. CAREER DYNAMICS

- A **mobilidade dos cientistas** é outro fator importante que oferece diversas oportunidades de carreira.
 - Ao examinar as mudanças no impacto associadas a cada movimento, quantificadas pelas citações, não foi encontrado nenhum aumento ou diminuição sistemática, nem mesmo quando os cientistas se mudaram para uma instituição de nível consideravelmente mais alto ou mais baixo
 - **Não é a instituição que cria o impacto; são os pesquisadores individuais que fazem uma instituição.**
- A **reputação** dos autores de artigos, medida pelo total de citações de sua produção anterior, aumenta acentuadamente o número de citações recebidas por esse artigo nos primeiros anos após a publicação. Após essa fase inicial, porém, o impacto depende da recepção do trabalho pela comunidade científica.
 - Para carreiras científicas produtivas, a reputação é menos um fator crítico para o sucesso do que talento, trabalho duro e relevância.
- Um modelo estocástico de evolução de impacto também indica que os avanços resultam de uma combinação da capacidade de um cientista e da sorte de escolher um problema com alto potencial.

IMPACT IN SCIENTIFIC CAREERS



A criatividade e a inovação dependem da idade ou do estágio da carreira?



(A) Registro de publicação de três prêmios Nobel em física.

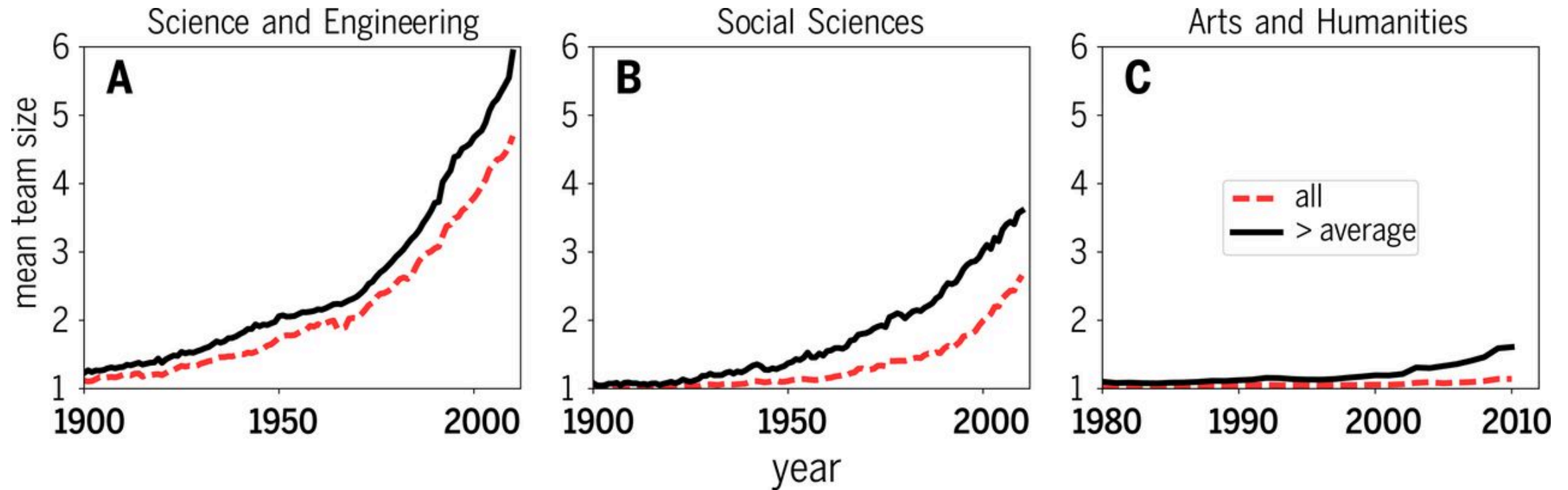
(B) Histograma da ocorrência do artigo de maior impacto na sequência de publicações de um cientista, calculado para 10.000 cientistas.

5. TEAM SCIENCE

- Durante as últimas décadas, a confiança no trabalho em equipe aumentou, representando uma mudança fundamental na forma como a ciência é feita.
 - Em 1955, as equipes de ciência e engenharia criaram o mesmo número de artigos que os autores individuais. No entanto, em 2013, a fração de artigos escritos por equipe aumentou para 90%.
 - Capacidade de apresentar mais combinações inovadoras de ideias ou produzir recursos que são posteriormente usados por outros.
 - Reunião de diferentes especialidades, combinando efetivamente conhecimento para levar a descobertas científicas.
 - O trabalho de grandes equipes gera, em média, mais citações em uma ampla variedade de domínios.
- Ter mais colaborações significa maior visibilidade por meio de um número maior de coautores, que provavelmente introduzirão o trabalho em suas redes, um impacto aprimorado que pode compensar parcialmente o fato de que o crédito dentro de uma equipe deve ser compartilhado com muitos colegas.
- A colaboração tem sido uma solução crítica, reunindo recursos para obter vantagem científica.

SIZE AND IMPACT OF TEAMS

O trabalho de alto impacto é mais provável de ser produzido por equipes grandes do que por pequenas.



5. TEAM SCIENCE

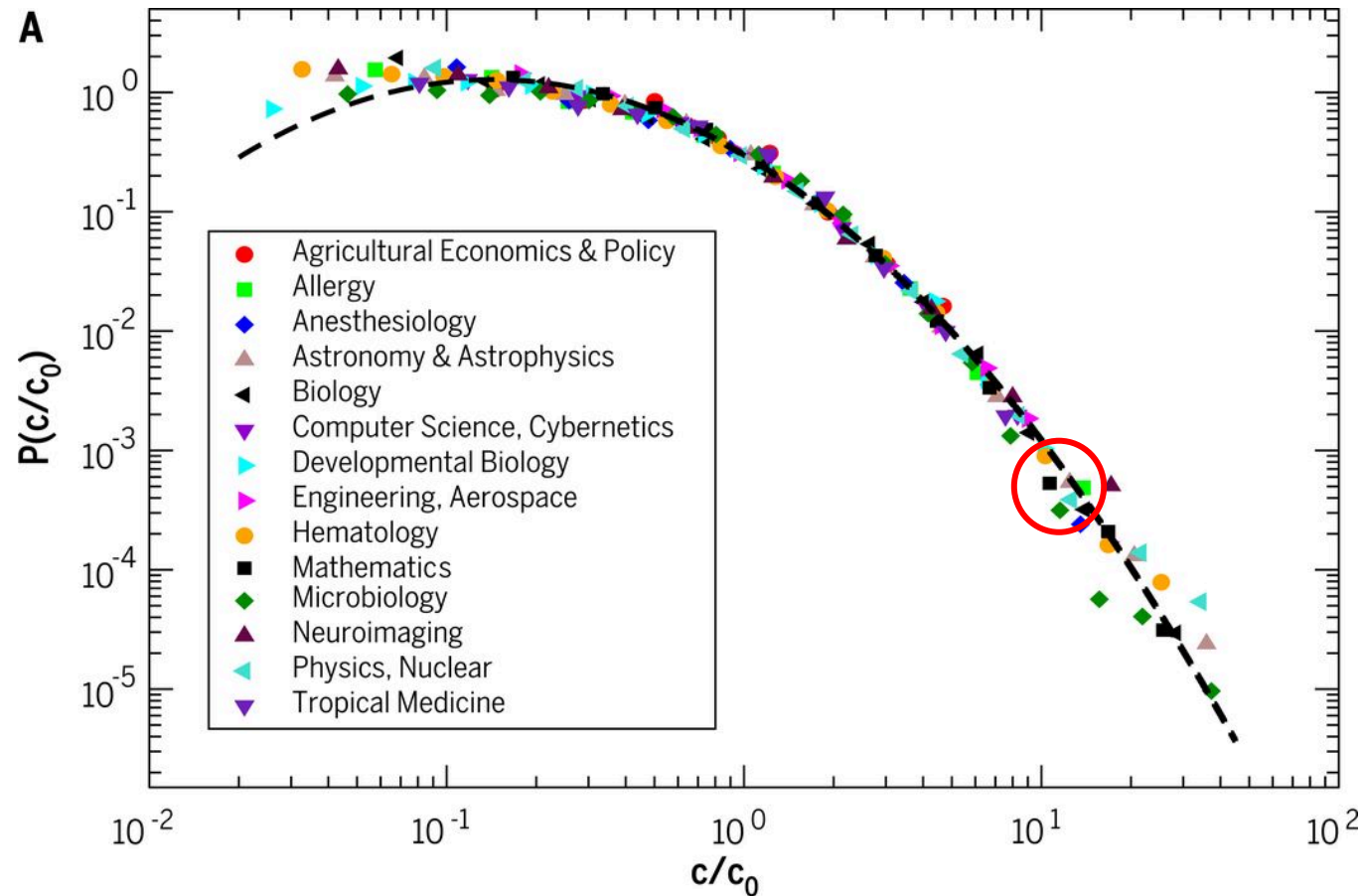
- **O tamanho é um determinante crucial das estratégias de sobrevivência da equipe:** equipes pequenas sobrevivem por mais tempo se mantiverem um núcleo estável, mas equipes maiores persistem por mais tempo se manifestarem um mecanismo de rotatividade de membros.
- Há um *trade-off* com o tamanho crescente que afeta o valor e o risco associados à “grande ciência” .
 - Embora possa ser possível resolver problemas maiores, o ônus da reprodutibilidade pode exigir a duplicação de esforços iniciais, que podem não ser viáveis prática ou economicamente.
- **Quem deveria ganhar mais crédito? Quem de fato recebe mais crédito?**
 - **Efeito Mateus** (atribuição errada de crédito): cientistas de *status* mais alto envolvidos em trabalho conjunto recebem crédito desproporcional por suas contribuições.
 - A alocação adequada de crédito individual para um trabalho colaborativo é difícil porque não podemos distinguir facilmente as contribuições individuais.

6. CITATION DYNAMICS

- A citação acadêmica continua sendo a unidade mensurável dominante de crédito na ciência. Dada a confiança da maioria das métricas de impacto nas citações, a dinâmica do acúmulo de citações foi examinada por gerações de estudiosos.
- **A distribuição de citações para artigos científicos é altamente distorcida:** muitos trabalhos nunca são citados, mas artigos seminais podem acumular 10.000 ou mais citações.
- A cauda da distribuição de citações segue uma lei de potências.
 - Geradas por meio de um processo de vantagem cumulativa - *preferential attachment*.
 - A probabilidade de citar um artigo aumenta com o número de citações que ele já recebeu.
- Um modelo preditivo assume que a **probabilidade de citação** de um artigo depende do número de citações anteriores, um fator de obsolescência e um parâmetro de aptidão (*fitness*).

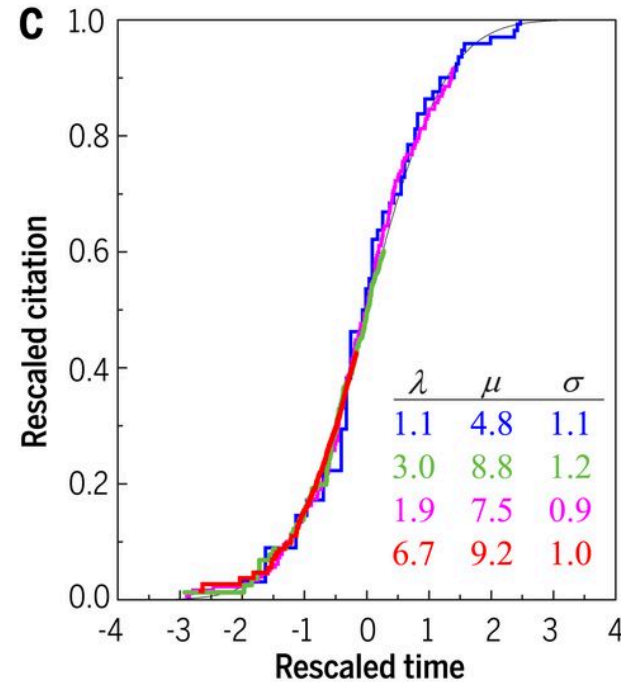
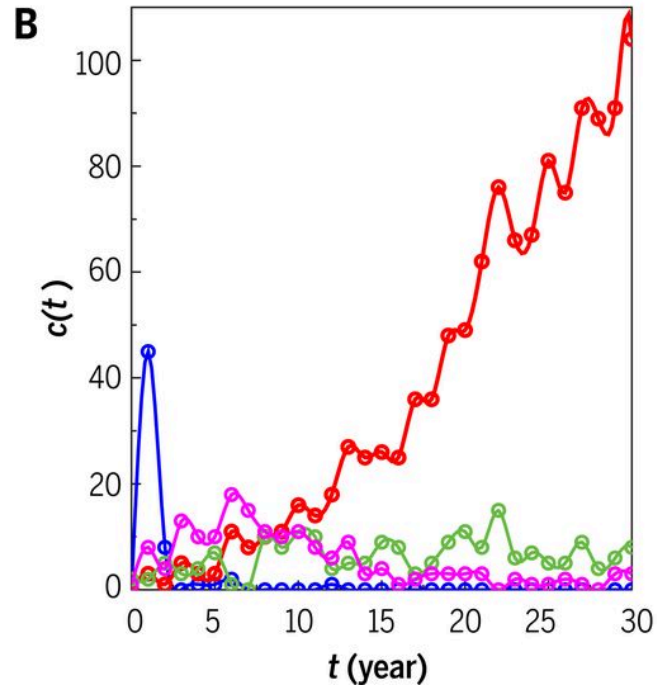
UNIVERSALITY IN CITATION DYNAMICS

Se o número de citações de um artigo é dividido pelo número médio de citações coletadas por trabalhos na mesma disciplina e ano, a distribuição da pontuação resultante é essencialmente indistinguível para todas as disciplinas.



(A) As distribuições de citações de artigos publicados na mesma disciplina e ano encontram-se na mesma curva para a maioria das disciplinas, se o número bruto de citações c de cada trabalho for dividido pelo número médio de citações c_0 sobre todos os trabalhos nessa disciplina e ano. A linha tracejada é um ajuste lognormal.

UNIVERSALITY IN CITATION DYNAMICS



(B) História de citação de quatro artigos publicados na *Physical Review* em 1964, selecionados por sua dinâmica distinta.

(C) As citações de um artigo individual i são determinados por três parâmetros: **fitness** λ_i , **imediaticidade** μ_i e **longevidade** σ_i .

7. OUTLOOK

- As diferenças entre as **perguntas**, os **dados** e as **habilidades** exigidas por cada disciplina sugerem que podemos obter mais informações a partir de estudos SciSci específicos de domínio que modelam e preveem oportunidades adaptadas às necessidades de cada campo.
- Para jovens cientistas, os resultados da SciSci oferecem *insights* úteis sobre padrões passados, ajudando a orientar futuras investigações dentro de suas disciplinas.
- A contribuição da SciSci é uma compreensão detalhada da estrutura relacional entre cientistas, instituições e ideias, um ponto de partida crucial que facilita a identificação de processos geradores fundamentais.
- A maioria das pesquisas SciSci se concentra em publicações como fontes primárias de dados, o que implica que os *insights* e descobertas são limitados a ideais bem-sucedidas o suficiente para merecer a publicação.

7. OUTLOOK

- No entanto, a maioria das tentativas científicas falha. Saber quando, por que e como uma ideia falha é essencial em nossas tentativas de entender e melhorar a ciência.
 - Esses estudos podem fornecer uma orientação significativa sobre a **crise de reprodutibilidade** e nos ajudar a explicar o **problema da gaveta de arquivos** (*file drawer problem*) – viés de publicação.
 - Eles também poderiam melhorar substancialmente nossa compreensão da imaginação humana, revelando o **fluxo total de atividade criativa**.
 - Pesquisadores podem desperdiçar tempo e esforço preciosos conduzindo pesquisas sobre tópicos que já foram bem pesquisados, mas que não foram relatados.
- A ciência pode ser melhorada através da ampliação do número e da variedade de indicadores de desempenho.
 - Métricas alternativas que cobrem a atividade da web e das mídias sociais e o impacto social. Também podemos incluir a informação (por exemplo, dados) que os cientistas compartilham com os concorrentes, a ajuda que oferecem aos seus pares e sua confiabilidade como revisores dos trabalhos de seus pares.

7. OUTLOOK

- A SciSci pode dar uma contribuição essencial ao fornecer modelos que ofereçam uma compreensão mais profunda dos mecanismos que governam os indicadores de desempenho na ciência.
 - Modelos de padrões empíricos observados quando **indicadores alternativos** (por exemplo, distribuições de downloads de artigos) são usados nos permitirão explorar sua relação com métricas baseadas em citações e reconhecer manipulações.
- A **integração de métricas baseadas em citações com indicadores alternativos** promoverão o pluralismo e possibilitarão novas dimensões de especialização produtiva, nas quais os cientistas podem ter sucesso de maneiras diferentes.
- Beneficiaria a ciência se a curiosidade, a criatividade e o intercâmbio intelectual fossem mais bem apreciados e incentivados no futuro. Uma abordagem mais pluralista poderia reduzir a duplicação e fazer a ciência florescer para a sociedade.

1. **Inovação e tradição:** deixadas à mostra, ideias verdadeiramente inovadoras e altamente interdisciplinares podem não alcançar o máximo impacto científico. Para aumentar seu impacto, novas ideias devem ser colocadas no contexto do conhecimento estabelecido.
2. **Persistência:** um cientista nunca é velho demais para fazer uma grande descoberta, contanto que ele continue produtivo.
3. **Colaboração:** a pesquisa está mudando para as equipes, portanto, envolver-se em colaboração é benéfico. Trabalhos de grandes equipes tendem a ter mais impacto.
4. **Crédito:** a maior parte do crédito irá para os coautores com o histórico mais consistente no domínio da publicação.
5. **Financiamento:** embora os painéis de revisão reconheçam a inovação, eles tendem a descontá-la. As agências de financiamento devem solicitar aos avaliadores que avaliem a inovação, não apenas o sucesso esperado.

DETECTABILIDADE DE GRUPOS ESTRATÉGICOS EM REDES SOCIAIS

PROJETOS EM ANDAMENTO

1. Estimar parâmetros (usar o método proposto pelo artigo Wang/Barabási) na rede real baixada da *Scopus*.
 - Serve para validar o artigo sobre os dados obtidos empiricamente.
 - Verificar que o método desenvolvido por eles se aplica na nossa rede real.
2. Dado que (1) já foi verificado, usar este método para a simulação semelhante à Tese do Walter.
 - Buscar os mesmos resultados da tese (análise do comportamento estratégico).
 - Mudar o processo de sorteio das citações conforme o utilizado no artigo do Wang/Barabási.
3. Validar por verossimilhança o método de geração de redes do TCC da Carol.
 - Com o modelo de sorteio das citações modificado (Wang/Barabási).
4. Usar algoritmo para identificação de comunidades, selecionar estratégicos e aplicar método para geração de rede por simulação (Tese do Walter) para buscar identificar comportamento estratégico.
 - Ideia apresentada no Workshop (Carol/Fernanda) – SINS 2017.