



# **Universidade de Brasília**

Faculdade de Administração, Contabilidade e Economia

Departamento de Economia

---

**Monografia**

## **Teoria da Informação, economia e a produção acadêmica**

Marcos Paulo Rodrigues Correia

2020

Brasília, DF

Marcos Paulo Rodrigues Correia

Matrícula: 160069157

## **Teoria da Informação, economia e a produção acadêmica**

Projeto de Trabalho de Conclusão de Curso apresentado na Faculdade de Administração, Contabilidade, Economia e Gestão de Políticas Públicas como requisito básico para a conclusão do curso de Ciências Econômicas

**Orientador: Dr. Bernardo Pinheiro Machado Mueller**

## **Sumário**

<b>1.</b>	<b>Introdução</b>	<b>4</b>
<b>2.</b>	<b>A Teoria da Informação</b>	<b>7</b>
<b>2.1.</b>	<b>Formalização do conceito</b>	<b>7</b>
<b>2.2.</b>	<b>Alternativas</b>	<b>13</b>
<b>2.3.</b>	<b>Informação Mútua</b>	<b>13</b>
<b>2.4.</b>	<b>Divergência entre distribuições e <i>KL</i></b>	<b>15</b>
<b>2.5.</b>	<b>Extensões e limitações</b>	<b>17</b>
<b>3.</b>	<b>O uso da teoria da informação</b>	<b>21</b>
<b>3.1.</b>	<b>Economia e teoria da informação</b>	<b>26</b>
<b>3.2.</b>	<b>Aplicações recentes</b>	<b>30</b>
<b>4.</b>	<b>Aplicação Empírica</b>	<b>35</b>
<b>4.1.</b>	<b>O Encontro Nacional de Economia</b>	<b>35</b>
<b>4.2.</b>	<b>Escopo da aplicação</b>	<b>36</b>
<b>4.3.</b>	<b>Resultados e discussões</b>	<b>42</b>
<b>5.</b>	<b>Conclusões</b>	<b>46</b>
	<b>Referências</b>	<b>47</b>

## 1. Introdução

Informação é um dos mais importantes conceitos em profusas áreas do conhecimento, da física à filosofia. John Archibald Wheeler, físico teórico, e Fred Dretske, filósofo da mente, esclarecem sua importância. Para Wheeler (1994) apud Dretske (1981), toda partícula, campo de força e o próprio espaço-tempo derivam suas funções a partir dos *bits*, referindo-se à unidade de medida da informação. Por esse motivo, Wheeler cria seu famoso lema: *it from bit*<sup>1</sup>. Para Dretske (1981), informação é a *commodity* capaz de produzir conhecimento caso as condições necessárias sejam respeitadas.

Ambos os autores se referem a um conceito específico sobre o que é a grandeza. Apesar de existirem outras definições, como a filosofia da informação focaliza-se em analisar (FLORIDI, 2002), a proposição de Claude Shannon (1948) marcou profundamente o saber científico dado sua interatividade com diferentes campos e sua elegância matemática. Chamada de Teoria da Informação (TI), trata-se de uma modelagem capaz de mensurar a quantidade de informação produzida por determinado processo, relacionando-a com o grau de aleatoriedade inerente a este.

O objetivo fundamental dessa abordagem reside em definir os limites e possibilidades da comunicação, de modo a entender os mecanismos pelos quais uma mensagem pode ser reproduzida perfeita ou aproximadamente de um ponto a outro (SHANNON, 1948). Embora mensagens tenham frequentemente valor semântico, a TI provém uma medida do conteúdo informativo, que é independente de sua significação.

---

<sup>1</sup> Refere-se ao fato de que todo o sistema físico tem como fonte um conceito misterioso, a informação.

À primeira vista, tamanho escopo pode parecer demasiadamente amplo ao passo que a formulação de uma teoria sobre informação longe de qualquer análise de sentido pode transparecer qualidade duvidosa. Como será perceptível, no entanto, são essas características que produzem a universalidade da teoria. Criptografia, telecomunicações, linguística, ciência da computação, teoria da probabilidade e de redes, estatística, matemática, física, neurociência e economia foram alguns dos campos impactados pela Teoria da Informação.

Especificamente, a relação entre TI e economia necessita algum cuidado, uma vez que a noção de informação na teoria econômica é polissêmica (GARROUSTE, 2001). Cabe destacar, nesse caso, que uma área tradicional e bem consolidada da literatura, a Economia Informacional (*Information Economics*), apresenta um conceito parcialmente independente daquele presente na modelagem de Shannon. Uma comparação entre essas abordagens será feita no terceiro capítulo desse trabalho.

Com essa ressalva, a TI possui extensa capacidade contributiva à pesquisa em economia. Por meio dela é possível mensurar incerteza, ruído, ordem e outras variáveis fundamentais para a análise econômica. Aplicações envolvendo Finanças e Teorias de Portfólio já angariam resultados promissores. Essa e outras contribuições da TI para a economia serão abordadas no terceiro capítulo.

Em especial, medida derivada da equação fundamental de Shannon, a divergência de Kullback-Leibler (KL) foi recentemente utilizada para medir inovação e transitoriedade, fator de grande interesse para a literatura. Um exame de recentes aplicações com KL será exposto, também, no terceiro capítulo deste trabalho e, no quarto, será feita uma aplicação própria com a metodologia para avaliar a inovação e transitoriedade das produções acadêmicas aceitas nos Encontros Nacionais de Economia, promovidos pela Associação Nacional dos Centros de Pós-graduação em Economia (ANPEC). O intuito é analisar quais áreas mais produziram inovação e quais destas conseguiram manter o debate econômico em linhas de pesquisa semelhantes, além de investigar possíveis vieses estruturais.

O presente documento visa produzir uma aplicação na área, bem como estabelecer uma revisão bibliográfica da Teoria da Informação. A monografia contará com cinco capítulos. No segundo, será realizada uma resenha da teoria da informação, abordando seus principais conceitos, extensões e limitações. O terceiro capítulo será destinado à exposição de aplicações em variados campos, brevemente, e na economia, enfaticamente. Nesse capítulo, também, será exposto algumas das mais interessantes contribuições recentes da teoria em sistemas sociais. No quarto, expor-se-á os resultados do estudo a respeito da novidade e transitoriedade na pesquisa econômica nacional. O último capítulo apresentará as conclusões do estudo e suas limitações.

## 2. A Teoria da Informação

### 2.1. Formalização do conceito

Informação e comunicação são conceitos interligados. No artigo “A Mathematical Theory of Communication”, Claude Shannon (1948) implicitamente articula a comunicação como qualquer tipo de processo no qual um mecanismo, biótico ou não, afeta outro, através da transmissão de informação. Como engenheiro de origem, seu interesse situava-se em desenvolver uma teoria geral da comunicação, com foco em aspectos da engenharia. Desejou, por exemplo, estabelecer o efeito do ruído sobre um canal de comunicação e entender até que ponto é possível reduzir a mensagem enviada sem perder sua estrutura. Weaver (WEAVER & SHANNON, 1949) acreditava que o estudo da comunicação possuía um problema central dividido em três níveis:

- **Nível A:** Com que precisão os símbolos (sinais) expressos na comunicação são transmitidos de um ponto a outro? (Problema técnico);
- **Nível B:** Quão preciso são os símbolos transmitidos em carregar o sentido desejado? (Problema semântico);
- **Nível C:** Quão efetivo é o sentido transmitido em produzir a resposta desejada ao destinatário? (Problema da efetividade).

A Teoria da Informação preocupa-se fundamentalmente com o problema técnico, de forma que o sentido da mensagem é irrelevante para sua formalização. Weaver (WEAVER & SHANNON, 1949) conclui que o tratamento ao Nível A limita e influencia os demais. Limita posto que a TI determina o limite máximo sobre o qual a informação presente em uma mensagem não pode exceder, da mesma forma que a renda de um agente limita seu consumo, ainda que não determine quais bens serão consumidos. Influencia uma vez que a própria modelagem dos

últimos dois níveis pode ser baseada na Teoria da Informação, exatamente como Dretske (1981) propôs desenvolver ao Nível B.

Para tratar o problema técnico, Shannon percebeu que a mensagem selecionada para transmissão é uma entre um conjunto de possíveis mensagens. A quantidade informativa de uma mensagem não estaria associada com seu conteúdo intrínseco, mas sim com o grau de surpresa que sua escolha produz. Isto é, informação é uma medida da incerteza ou, equivalentemente, da liberdade de escolha em determinado contexto (WEAVER & SHANNON, 1949). Um exemplo pode ajudar a entender essa forma pouco convencional de tratar a informação.

Suponha uma situação fictícia onde é necessário escolher um entre oito economistas para a vaga como professor visitante, cada um com mesma probabilidade de ser selecionado. Um membro do comitê informa posteriormente que, desses oito, somente dois respeitam totalmente as condições exigidas no processo. Após o comunicado, a incerteza quanto à seleção do profissional foi reduzida das oito possibilidades iniciais para duas e, assim, o membro do comitê transmitiu informação.

Na atual situação elementar onde há de se escolher entre duas possibilidades, é arbitrariamente dito que a informação associada com esse contexto é a unidade (WEAVER & SHANNON, 1949), medida pelos *bits*. Note que é a situação de escolha que possui uma unidade de informação e não suas possibilidades particulares, de maneira que a opção por qualquer um dos dois candidatos gera a mesma quantidade de informação. É por esse motivo que a TI é independente de noções de sentido.

A mensuração da quantidade<sup>2</sup> de informação é direta, realizada a partir da representação de processos de seleção como variáveis aleatórias. Seja  $X$  uma variável aleatória de distribuição uniforme discreta e  $n$  a quantidade de elementos,

---

<sup>2</sup> É mais correto pensar em taxa de informação ante quantidade, uma vez que a fonte de informação pode a produzir continuamente. Para efeito de exposição da teoria, tratar de quantidade informativa não causa perda de sua generalidade.



neste caso, as possibilidades. Então, define-se  $I(X)$  como a quantidade de informação produzida em  $X$ , de modo que:

$$I(X) = n \quad (2.1)$$

onde  $\log_2$  é o logaritmo na base 2. O uso de logaritmos é conveniente já que é matematicamente fácil realizar operações, além de representar adequadamente o modo pelo qual parâmetros de interesse crescem, como o tempo e os transistores em um sistema computacional (SHANNON, 1948). Além disso, a escolha da base do logaritmo corresponde com a escolha da unidade de medida da informação, que se mostrou mais intuitiva como a base binária.  $I(X)$  é chamado de quantidade esperada de informação em  $X$  e de entropia da distribuição uniforme.

Se as alternativas não são igualmente prováveis, (2.1) não pode ser utilizada. Suponha, por exemplo, que um dos economistas restantes no processo possui probabilidade de 80% de ser selecionado, dado que já possuiu experiência como professor e produziu artigos em revistas prestigiadas. Nesse caso, sua efetivação produzirá, em média, menos informação pois menos incerteza estará associado ao processo de escolha. Em outras palavras, menos surpreendente seria a sua escolha para o cargo. De modo geral, circunstâncias em que uma alternativa possui elevada probabilidade de ocorrer produzirão pouca informação. No limite, quando uma alternativa qualquer ocorre com certeza, o processo não produz informação.

Assim, a quantidade de informação produzida em um processo generalista é representada da seguinte forma. Seja  $X$  uma variável aleatória discreta e  $n$  a quantidade de elementos no espaço amostral  $\Omega = x_1, x_2, \dots, x_n$ . Define-se a informação presente no processo  $X$  como:

$$I(X) = \sum_{i=1}^n p(x_i) \log_2 \left( \frac{1}{p(x_i)} \right) \quad (2.2)$$

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (2.3)$$

no qual  $p(x_i)$  é a probabilidade do evento  $x_i$  ser o escolhido. O termo  $-\log_2 p(x_i)$  determina a quantidade de informação que cada escolha individualmente produz.<sup>3</sup> Uma vez que o interesse é a informação total na situação de escolha, pondera-se esse valor pela sua probabilidade de ocorrer,  $p(x_i)$ , e soma-se todas as possibilidades. O resultado é a estimativa da quantidade esperada de informação. No caso em que um dos economistas possui 80% de chance de ser selecionado, denotado por  $X'$ , a decisão produz:

$$H(X') = -(0,8 \log_2 0,8 + 0,2 \log_2 0,2) = 0.7219 \text{ bit} \quad (2.4)$$

ao passo que, no caso inicial com ambos desfrutando da mesma chance de serem selecionados ( $X''$ ), temos:

$$H(X'') = -(0,5 \log_2 0,5 + 0,5 \log_2 0,5) = \log_2 2 = 1 \text{ bit} \quad (2.5)$$

Note que (2.5) equivale a equação (2.1) quando  $n = 2$ , situação que produziu mais informação. Por estranho que pareça, a quantidade de informação cresce à medida que o processo se torna mais aleatório (GLEICK, 2011).

As equações (2.2) e (2.3) são matematicamente equivalentes, contudo, o termo  $I(X)$  foi substituído por  $H(X)$  para se referir a entropia de Shannon que, por sua vez, usou H para indicar o teorema H de Boltzman. Entropia é uma medida de desordem, incerteza ou volatilidade associada a uma variável aleatória (MAASUOMI, 1993). Maasuomi (1993) revela que existe uma relação tenaz, extensamente debatida, entre  $H(X)$  e a entropia física apresentada na segunda lei da termodinâmica<sup>4</sup>. Curiosamente, a sugestão de como nomear (2.3) proveio do famoso matemático John Von Neumann, conforme Myron Tribus (TRIBUS & MCIRVINE, 1971, página 180) recordou em conversa com Shannon:

---

<sup>3</sup> Há uma outra forma de interpretar  $-\log_2 p_{xi}$ . O termo representa a menor quantidade de bits (ou perguntas binárias) capazes de codificar eficientemente um evento  $x_i$  de probabilidade  $p_{xi}$ . Assim, a entropia de Shannon representa a quantidade esperada de bits para representar o evento que de fato ocorre. Em aplicações, quando  $-\log_2 p_{xi}$  não é inteiro, toma-se o menor inteiro  $z$  tal que  $z > -\log_2 p_{xi}$ . Por exemplo, para  $p_{xi}=0,2$  temos:  $-\log_2 0,2 \cong 2,322$ . Isso significa que são necessários, no mínimo, 3 bits para descrever eficientemente  $x_i$  em linguagem binária.

<sup>4</sup> Entropia de Gibbs, utilizada amplamente na física, é equivalente a entropia de Shannon com a adição da constante de Boltzmann,  $k_b$ :  $S = -k_b \sum p_i \log p_i$ . Pela facilidade de manipulação, a entropia de Shannon pode ser convertida na de Gibbs através da mudança de base logarítmica.

My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.'

Como a própria relação com a física sugere, a estrutura matemática da entropia de Shannon não é mero acaso. Na realidade, qualquer função não negativa,  $g(\cdot)$ , aplicada à função de probabilidade de uma variável aleatória  $X: \Omega \rightarrow R$ , que satisfaz:

$$g(p(x_i)) = \begin{cases} 0, & \text{caso } p(x_i) = 1 \\ \infty, & \text{caso } p(x_i) \rightarrow 0 \end{cases}, \text{ para } \forall x_i \in \Omega \quad (2.6)$$

é uma candidata razoável para mensurar informação (MAASUOMI, 1993). A fim de encontrar a ideal, Claude Shannon definiu propriedades de interesse que foram, posteriormente, transformadas nos seguintes axiomas (DEDEO, 2016):

1. **Continuidade:**  $g(p(x_1), \dots, p(x_n))$  deve ser contínua nos possíveis  $p(x_i)$ ;
2. **Simetria:**  $g(p(x_1), \dots, p(x_n))$  é uma função simétrica nos possíveis  $p(x_i)$ . Assim, a reordenação das probabilidades de cada evento não produz mudança na quantidade de informação do processo. A simetria reitera sua não relação com noções de sentido;
3. **Condição de Máxima Informação:**  $g(p(x_1), \dots, p(x_n))$  tem valor máximo quando  $p(x_i) = \frac{1}{n}$ , para  $\forall i \leq n$ ;

4. **Coarse-Graining ou recursão:** Para todo  $0 \leq \lambda < 1$  vale que:

$$g(p(x_1), p(x_2), \dots, \lambda p(x_n), (1 - \lambda)p(x_n)) = g(p(x_1), p(x_2), \dots, p(x_n)) + p(x_n)g(\lambda, 1) \quad (2.7)$$

É possível provar que a única função  $g$  que satisfaz as premissas acima é da forma:<sup>5</sup>

---

<sup>5</sup> Na abordagem axiomática da entropia estatística, existem outras propriedades que podem ser utilizadas para provar que a entropia de Shannon é a ideal para uma grande variedade de contextos. Os quatro axiomas expressos são os mais tradicionais para tal prova.

$$g(.) = -K \sum_{i=1}^n p(x_i) \log \log p(x_i) \quad (2.8)$$

onde  $K$  é uma constante positiva alterada pela determinação da base logarítmica.

O axioma de continuidade encontra-se presente em diversas representações matemáticas de fenômenos naturais e sociais, por isso sua necessidade é bastante aceitável. A simetria e a condição de máxima informação garantem a abstração conceitual desejada ao tratar um conceito de escopo tão amplo. Já o interesse no axioma de recursão necessita de uma explicação mais cuidadosa.

*Coarse-graining* trata de como agrupar a informação. Ao optar por transmitir a frase “o carro que passou era elétrico” ao invés de “o carro que passou era elétrico e de cor amarela” ocorre uma deliberada redução da quantidade de informação transmitida e, assim, aplica-se o princípio de recursão (DEDEO, 2016). A informação perdida representa o segundo termo do lado direito de (2.7), que pode ser útil ou não, a depender do contexto. Nesse sentido, o quarto axioma determina que informação pode ser compactada e sua perda informativa mensurada.

Em comparação, a macroeconomia utiliza o mesmo princípio ao propor que variáveis econômicas agregadas podem explicar o comportamento da economia como um todo. De fato, o próprio processo científico resulta na criação de modelos simplificadores da realidade para explicar e prever seu comportamento sistêmico. Nos dois casos, parte da informação é retirada, assumindo que não é preciso sua integralidade para que alguma compreensão ocorra. Quando a exposição detalhada e perfeita de um sistema não pode ser realizada, o *Coarse-graining* é a solução.

Assim, o respeito aos quatro axiomas garante a existência de propriedades especiais para a entropia. A medida possui comportamento de árvore lógica, o que possibilita aplicação para muitos problemas computacionais. Além disso, a noção de entropia na Teoria da Informação é o melhor mecanismo de compreensão de dados (COVER & THOMAS, 1991).

## 2.2. Alternativas

Como conclui DeDeo (2016), a entropia de Shannon, age como um indicador de diversidade. No campo econômico, o índice de Herfindahl-Hirschman foi criado com o intuito de mensurar a concentração em um mercado e, assim, a competitividade e diversidade no setor (RHOADES, 1993). O índice,  $D$ , é medido como:

$$D(S) = \sum_{i=1}^N s_i^2 \quad (2.9)$$

no qual  $S$  representa o conjunto com o *market share* das  $N$  firmas presentes nesse mercado e  $s_i$  o *market share* de cada firma. Uma vez que cada  $s_i$  é sempre menor ou igual a unidade e nunca menor que zero, é possível entender a medida como a probabilidade de se selecionar uma empresa específica no conjunto de todas as firmas da indústria. Por isso, o indicador pode ser um possível substituinte para a entropia de Shannon.

O índice de Herfindahl-Hirschman é muitas vezes utilizado para políticas antitrustes de maneira que tem grande importância empírica. No entanto, o indicador não respeita o axioma de *Coarse-graining*, que possibilita a entropia possuir características atraentes. Em suma, nenhuma medida de diversidade em outros campos científicos satisfaz os quatro axiomas da entropia e, portanto, esta é a melhor medida para essa mensuração (DEDEO, 2016).



### 2.3. — Informação Mútua

Definida a informação presente no emissor e suas propriedades, resta resolver o problema técnico, qual seja, com que precisão os símbolos são transmitidos de um ponto ao outro. Shannon (1949) ponderou que a transmissão de informação ocorre por um canal de comunicação, local onde a mensagem produzida no emissor é codificada em símbolos/sinais e deve ser decodificada pelo receptor. Se, por exemplo, o comitê decide enviar uma carta com o resultado do processo seletivo, a mensagem é codificada fisicamente em caracteres alfabéticos e posteriormente decodificada no cérebro de cada candidato. Dessa maneira, a comunicação depende da capacidade do canal para transmitir informação de um ponto a outro.

A informação que o destinatário recebe também pode ser manipulada como uma variável aleatória. Seja  $Y$  essa variável, com  $m$  elementos. Caso o canal comunicativo não tenha influência de ruído, então  $H(X) = H(Y)$ . Na maior parte dos casos, contudo, o ruído tem efeito considerável sobre a transmissão de informação, circunstâncias em que as entropias são apenas parcialmente correlacionadas. Observando pelo ponto de vista do destinatário, a questão de interesse é determinar o quanto se desconhece sobre  $X$  sabendo que  $Y$  ocorreu. Para isso, define-se a entropia condicional de  $X$  dado  $Y$  como:

$$H(X|Y) = - \sum_{i,j=1}^{n,m} p(x_i \cap y_j) \log_2 p(y_j) \quad (2.10)$$

$H(X|Y)$  determina a incerteza média sobre  $X$  após observar  $Y$ . Em outras palavras, esse termo representa a quantidade de informação que é perdida da fonte após a transmissão para o destinatário. A partir dessa lógica, pode se definir a informação mútua entre duas variáveis,  $I(X; Y)$ , como a quantidade de informação obtida sobre uma destas conhecendo-se a outra (GHAHRAMANI, 2006):

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (2.11)$$

Informação mútua é simétrica entre as duas variáveis. Quando o canal utilizado não apresenta ruído toda a informação produzida em  $X$  é transmitida, ou seja,  $H(X, Y) = H(X)$ . Basta conhecer  $H(X)$  para saber exatamente o que ocorre em  $H(Y)$  e vice-versa. No outro extremo, quando as variáveis  $X$  e  $Y$  são completamente independentes, não há informação mútua, isto é,  $H(X) = H(X|Y)$ . Isso significa que conhecer ou não  $Y$  é irrelevante pois nada que ocorre em  $X$  é conhecido por  $Y$ .

Shannon (1949) utilizou o conceito de informação mútua para definir uma propriedade, a capacidade do canal. Essa é a maior taxa de informação (em *bits* por segundo) que pode ser transmitida em um dado canal de comunicação com probabilidade de erros arbitrariamente baixa. A capacidade do canal,  $C$ , é medida como:

$$C = (H(X) - H(Y)) \quad (2.12)$$

onde maximiza-se com respeito a todas as possíveis fontes de informação  $H(X)$ , que podem ser entradas no canal. Sempre que a entropia da fonte supera a capacidade do canal, erros na transmissão ocorrem frequentemente. Com esse aparato teórico, Shannon estabeleceu sua contribuição para o estudo da comunicação, determinando resposta robusta ao problema técnico.

O autor, além disso, mostrou uma possível solução para situações de excesso de informação no canal: basta aumentar a ambiguidade (redundância) da mensagem. Aumentar a ambiguidade é análogo a tornar a mensagem mais repetitiva. Assim, mesmo que a informação transmitida sofra bastante alteração, o que alcança  $Y$  eventualmente recuperará a informação em  $X$ . Tal princípio é bastante aplicável em campos de estudo que necessitam de rigor na informação transmitida, como a criptografia.

#### **2.4. Divergência entre distribuições e KL**

Uma vez que TI trabalha com manipulação de variáveis aleatórias, é quase imediato perceber a relação entre o campo e a Estatística. Um dos ramos mais influenciados por TI é a Inferência Estatística, que pode ser entendida como o campo focado em mensurar e julgar afinidades e diferenças entre distribuições (MAASUOMI, 1993). Há várias bases metodológicas para testar hipóteses, recurso comum na Inferência, conquanto TI fornece um grupo poderoso de testes a partir da entropia das distribuições analisadas. Entre esses testes, a Divergência de Kullback-Leiber (KL) é geralmente o critério de escolha em função de suas características convenientes, como respeitar o axioma de *coarse-graining* (MAASUOMI, 1993).

A divergência de Kullback-Leiber é uma medida da diferença entre duas distribuições de probabilidade, de modo que serve como medida de qualidade do ajuste entre essas distribuições (EVREN & TUNA, 2012). Na linguagem informacional, KL mensura o ganho informacional ao se utilizar um modelo distributivo aproximado em relação a distribuição efetiva. A fórmula para KL requer apenas uma leve modificação em relação a entropia de Shannon (2.3). Seja  $q(x)$  a função de probabilidade relativa à distribuição aproximada de  $x$  e  $p(x)$  a distribuição verificada. Então:

$$KL(q) = \sum_{i=1}^n p(x_i) (\log_2 p(x_i) - \log_2 q(x_i)) \quad (2.13)$$

$$KL(q) = \sum_{i=1}^n p(x_i) \log_2 \frac{p(x_i)}{q(x_i)} \quad (2.14)$$

Pela desigualdade de Jensen,  $KL(q) \geq 0$  e  $KL(q) = 0$  quando  $p(x) = q(x)$  (MAASUOMI, 1993). Note que, apesar da divergência de Kullback-Leiber ser escrita como  $p$  condicionado a  $q$ , a equação (2.14) difere da entropia condicional definida em (2.10). O termo no logaritmo é uma razão entre as distribuições, diferente da probabilidade condicional em (2.10), e pondera-se apenas pela probabilidade  $p(x_i)$ . A medida determina, em essência, a esperança da diferença logarítmica entre a probabilidade dos dados na distribuição original e na distribuição aproximada. Dessa forma:



$$KL(q) = E[p(x) - q(x)] \quad (2.15)$$

A divergência de Kullback-Leiber não é simétrica em relação as distribuições já que não respeita a desigualdade triangular. Em função disso, fala-se em uma divergência e não distância entre duas distribuições. Quando simetria é uma propriedade de interesse, utiliza-se a Distância de Jensen-Shannon, criada diretamente de KL.

A interpretação da grandeza é simples. Quanto maior for KL, mais informação é ganha ao aproximar  $p(x)$  por  $q(x)$ . Comparar  $KL(q)$  com, por exemplo,  $KL(h)$ , a divergência de outra distribuição aproximada  $h(x)$  com  $p(x)$ , permite avaliar qual dos dois modelos distributivos melhor se adequa aos dados. Como o ganho de informação em KL representa aumento de incerteza sobre  $p(x)$ , seleciona-se a distribuição aproximada de menor KL.

A grande vantagem de testes informacionais, como KL, é que eles não dependem de dimensão ou parametrização das distribuições analisadas (MAASUOMI, 1993), permitindo uso geral em aplicações econométricas. De fato, alguns dos principais testes para seleção de modelos são derivados de KL, como *General information criterion* (GIC), *Akaike's information criterion* (AIC) e *Bayesian information criterion* (BIC) (EVREN & TUNA, 2012). Testes de má especificação de modelo também possuem relação com a divergência, desde que utilize-se *Fischer Information*<sup>6</sup> para sua formalização.

Um aspecto fundamental de KL diz respeito a sua interpretação a partir do raciocínio bayesiano. Pode-se descrever a divergência de Kullback-Leiber em relação a “surpresa” (“surpresa Bayesiana”) de uma distribuição  $p(x)$  dado que um agente esperava  $q(x)$  (DEDEO, 2016). A surpresa seria uma medida da inovação presente na distribuição verdadeira,  $p(x)$ , em relação a distribuição esperada  $q(x)$ . Nessa abordagem, KL representa a interpretação epistêmica de um agente testando suas crenças,  $q(x)$ , e comparando-as com novas evidências,

---

<sup>6</sup> Pode ser mostrado que *Fischer Information* é a derivada segunda da Divergência de Kullback-Leiber entre a distribuição aproximada  $f(x;\theta)$  e a distribuição verdadeira  $f(x;0)$ , onde  $\theta$  é um parâmetro da distribuição.

$p(x)$ . Essa formulação tem sido de enorme aplicabilidade para estudos sobre inovação, que serão expostos no próximo capítulo deste trabalho.

## 2.5. Extensões e limitações

Apresentada até aqui, TI aplica-se a fontes discretas de informação, isto é, fontes cujos símbolos e mensagens podem ser representadas através de variáveis aleatórias discretas. Para fontes contínuas a fórmula é estendida a partir de integrais e funções de densidade de probabilidade, não obstante alguns cuidados serem tomados a fim de especificar o nível de precisão com a qual a variável será representada (GHAHRAMANI, 2006). Em qualquer caso, o resultado que Shannon provou sobre a melhor forma funcional para a mensuração da informação, em (2.8), só é válido se a fonte geradora for estocástica<sup>7</sup> e ergódica<sup>8</sup>.

É razoável formar alguma dúvida sobre a universalidade da TI ao constatar-se que a evolução da vida e dos sistemas sociais são puramente processos não ergódicos. Essa limitação foi e tem sido tratada de diferentes formas. Em primeiro lugar, Shannon (1948) construiu exemplos de processos não ergódicos divisíveis em um número finito de subprocessos estatisticamente homogêneos (ergódicos) e os cunhou de fontes “misturadas”. Nesses casos, toda vez que o processo inicia, um de seus subprocessos é escolhido, de acordo com uma probabilidade, e segue indefinidamente nesse estado. Assim, a entropia de fontes “misturadas” foi definida como a esperança da entropia dos subprocessos puros. Solução engenhosa, mas limitada.

Em segundo lugar, deve-se levar em conta o contexto de aplicação. Desde o artigo original, a linguagem tem sido importante objeto de análise para a TI,

---

<sup>7</sup> Informalmente, um processo estocástico é uma família de variáveis aleatórias que evoluem no tempo de acordo com certas probabilidades. Muitos eventos podem ser modelados como estocásticos.

<sup>8</sup> Informalmente, um processo ergódico é um tipo de processo no qual amostras razoavelmente largas tendem a representar o processo com um todo (WEAVER & SHANNON, 1949). Muitos eventos de interesse na economia e ciências sociais não são ergódicos.

influenciando campos da linguística como *Natural Language Processing (NLP)*. Mesmo que a linguagem, em si, tenha estrutura não ergódica, é possível a modelar por meio de cadeias de Markov (MANNING & SCHUTZE, 1999). Para isso, basta assumir que a probabilidade de uma palavra (ou letra) ser escolhida em um texto depende das  $k$  palavras (ou letras) anteriores, o que produz uma aproximação de Markov de ordem  $k$ :

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_{n-k} = x_{n-k})$$

(2.16)

Com essa modelagem, aplica-se a entropia de Shannon na matriz de transição construída através das probabilidades em (2.16). Algoritmos utilizados para modelar textos utilizam-se de cadeias de Markov para sua formalização e permitem, por isso, aplicações de TI.

Uma outra solução consiste em definir uma função de entropia capaz de mensurar processos não ergódicos e, em especial, sistemas complexos.<sup>9</sup> A iniciativa foi possível por meio da alteração dos axiomas anteriores, de maneira a melhor representar o comportamento de tais sistemas. Thurner, Corominas-Murtra e Hanel (2017) produziram alguns resultados nesse campo, encontrando diferentes formas funcionais para a entropia a partir de generalizações dos seguintes tipos de processos: Urna de Polya, *Sample Space Reducing (SSR)* e Multinomial Mixture (MMP).<sup>10</sup>

A modelagem da informação em processos não ergódicos é relativamente recente e ainda restam formalizações para sistemas complexos mais abrangentes, cujo comportamento é *self-reinforcing* e *path-dependent* simultaneamente (THURNER et al, 2017). Felizmente, os resultados no campo são promissores e já permitem mensurar a informação em distribuições não convencionais como *Power Law* e *Zipf's law*.

---

<sup>9</sup> Sistemas complexos são sistemas com grande número de agentes, normalmente heterogêneos, interagindo entre si, exibindo comportamento emergente e auto organizado.

<sup>10</sup> Urna de Polya é um exemplo de processo *self-reinforcing*, característica comum em sistemas complexos. SSR é um exemplo de processo *path-dependent*, característica também comum em sistemas complexos. MMP é um exemplo de processo estocástico composto, característicos de processos estocásticos que contém saltos.

Por fim, uma abordagem moderna à informação, conhecida como Teoria Algorítmica da Informação (TAI), pode mensurar a grandeza em processos variados, sejam estocásticos e ergódicos ou não. Sua utilidade reside na formulação de uma medida não probabilística, adequada a contextos cuja caracterização do espaço amostral é inacessível. Novamente, alguma formalidade matemática será necessária. Seja  $x$  o objeto (processo) de interesse expresso como uma cadeia finita de caracteres binários (o que implica ser um objeto computável),  $l(x)$  seu comprimento e  $U$  um computador universal.<sup>11</sup> Seja, também,  $U(p)$  o resultado em  $U$  quando embutido um programa  $p$ . Então define-se a complexidade de Kolmogorov (*Kolmogorov Complexity*),  $K$ , para uma cadeia binária  $x$  com respeito a  $U$ , como:

$$K_U(x) = l(p) \tag{2.17}$$

$K$  é o menor entre todos os possíveis programas que se encerram quando produzem  $x$  (COVER e THOMAS, 1991). A complexidade de Kolmogorov determina a informação de  $x$ . Conforme Gleick (2011) expõe, o interesse em relacionar a informação de um objeto com algoritmos foi originado de modo independente por Solomonoff, Kolmogorov e Chaitin:

Solomonoff, Kolmogorov, and Chaitin tackled three different problems and came up with the same answer. Solomonoff was interested in inductive inference: given a sequence of observations, how can one make the best predictions about what will come next? Kolmogorov was looking for a mathematical definition of randomness: what does it mean to say that one sequence is more random than another, when they have the same probability of emerging from a series of coin flips? And Chaitin was trying to find a deep path into Gödel incompleteness by way of Turing and Shannon—as he said later, “putting Shannon’s information theory and Turing’s computability theory into a cocktail shaker and shaking vigorously.” They all arrived at minimal program size. And they all ended up talking about complexity.

---

<sup>11</sup> Computador universal é um computador que pode simular as ações de outros computadores (COVER e THOMAS, 1991). Computadores digitais são universais e, em particular, podem simular uma Máquina de Turing Universal, que é um dispositivo teórico utilizado para representar as operações lógicas básicas de um computador (WIKIPEDIA). Church (1936) prova que todos os modelos computacionais suficientemente complexos são equivalentes, pois só computam uma mesma família de programas. Esse resultado é diretamente relacionado ao teorema de Kolmogorov.

O resultado central que possibilitou todas as investigações posteriores foi o teorema da universalidade de Kolmogorov (LEVIN, 1971). Nele é provado que, dado uma máquina universal  $U$ , para qualquer outra máquina  $A$ , existe uma constante  $c_A$  tal que:

$$K_U(x) \leq K_A(x) + c_A \quad (2.18)$$

em que  $x$  é uma cadeia binária qualquer e  $c_A$  uma constante independente de tal cadeia (COVER e THOMAS, 1991). À medida que  $x$  se torna mais complexo,  $c_A$  se torna assintoticamente pequeno, ou seja, a complexidade de Kolmogorov é uma característica fundamental de  $x$ . O resultado prova que é possível determinar a quantidade informativa, suficientemente invariante, de qualquer objeto computável (LEVIN, 1971).

Com essas soluções, é possível argumentar que a TI alcançou algum grau de universalidade. Embora a Teoria Algorítmica da Informação detenha um sentido ainda mais profundo que a proposta original, aplicações da entropia de Shannon, bem como outros aspectos de seu trabalho, são bastante poderosos e mais comuns ao campo econômico. Por isso, o presente trabalho focaliza na análise de TI, ao passo que referências à TAI só serão realizadas quando explicitamente expressas.



### **3. O uso da teoria da informação**

Após o artigo de Shannon (1948), a nova modelagem rapidamente se espalhou. Cibernética, computação, linguagem, cognição, estatística além de várias outras áreas de conhecimento utilizaram a Teoria da Informação em seus objetos de interesse. No entanto, durante tal processo de sincretismo, exageros e má interpretações foram recorrentes.

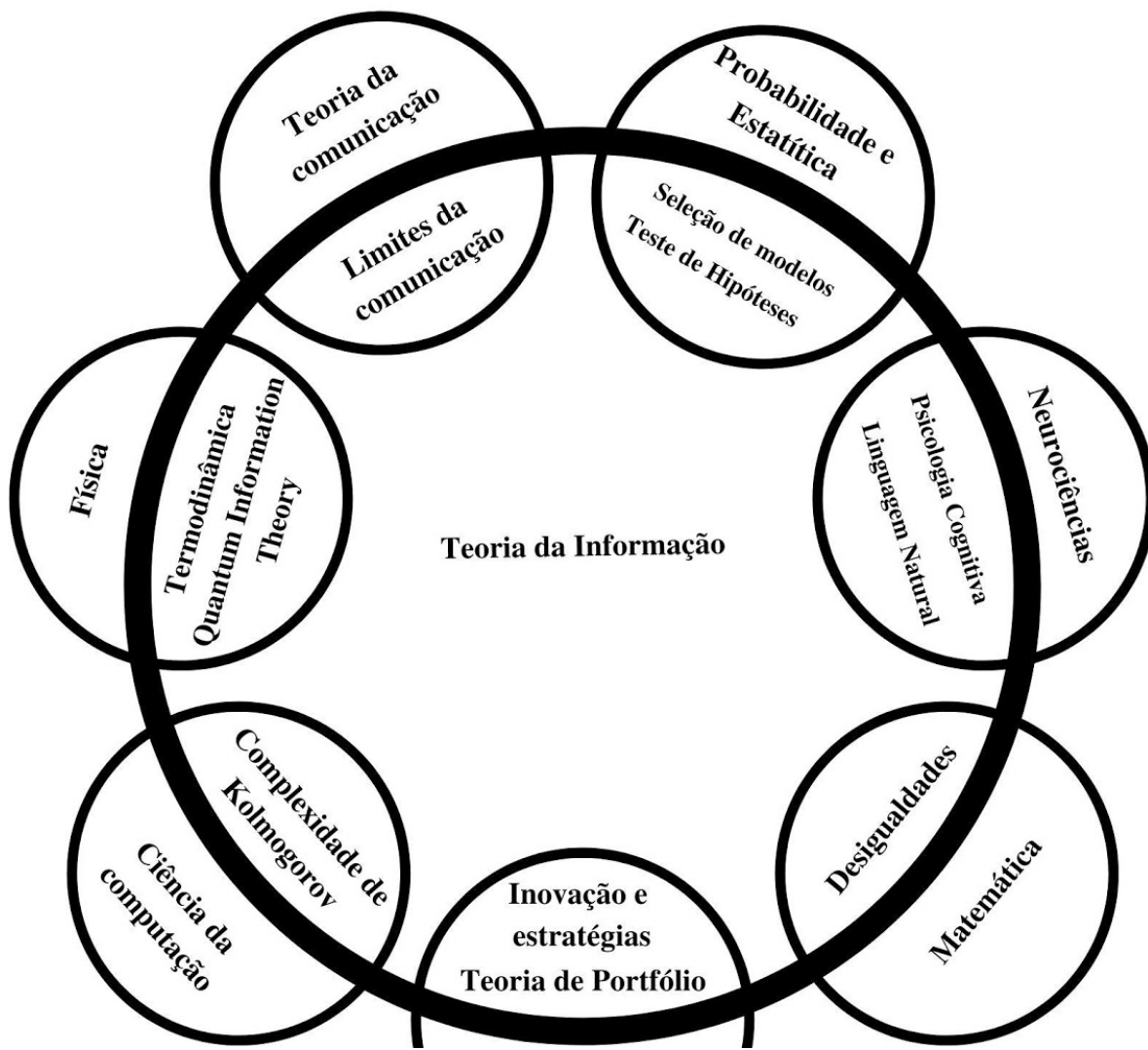
O próprio matemático alertou para o uso indiscriminado da teoria sem um respaldo teórico efetivo, em que algumas definições eram traduzidas para diferentes domínios e entendidas como verdades a partir de fatos matemáticos e não de testes para sua devida validação (SHANNON, 1956). Se, a título de exemplo, um neurocientista afirmasse que o cérebro humano possui comportamento de um decodificador ideal, seria necessário um experimento para testar tal hipótese enquanto a prova formal de sua possibilidade não seria evidência suficiente.

DeDeo (2016), felizmente, argumenta que a inquietação de Shannon foi solucionada e, aos poucos, TI foi compreendida em sua extensão e limitação. A

Revolução Digital teve grande papel nisso, posto que boa parte dos experimentos necessitavam de um volume de dados dificilmente processáveis para computadores comuns à década de 60 (AFTAB et al, 2001). Testes para avaliar a adaptação de TI a diferentes campos se tornaram cada vez mais fáceis ao mesmo tempo que o desenvolvimento teórico crescia com vigor.

Nesse meio, algumas ciências se destacaram pelo benefício adquirido com a adoção de Teoria da Informação. Cabe destacar sete delas: teoria da comunicação; probabilidade e estatística; neurociências; matemática; física; ciência da computação e economia. Cada uma possui contribuições diferentes e vale ressaltá-las individualmente. No caso da economia, um tratamento específico e cuidadoso será realizado por se tratar do foco deste trabalho.

**Figura 1 - Relação entre Teoria da Informação e outros Campos\***



Em Teoria da Comunicação, a modelagem de Shannon produziu uma robusta solução ao problema técnico da comunicação. Trabalhos recentes com os aspectos comunicativos concentraram-se no estudo de redes, especialmente a análise de como é transmitida informação em um sistema de *network* e de que forma os nós cooperam para essa transmissão (XIE e KUMAR, 2004).

Quanto à Estatística e Probabilidade, conforme anteriormente exposto, a Teoria da Informação possui interação profunda. Além das aplicações em testes de hipóteses e medidas para qualidade de ajuste, há um vasto número de pesquisas conjuntas. Uma coleção dessas aplicações expandiu noções sobre o comportamento de longas sequências de variáveis aleatórias, facilitando estimação de eventos raros (COVER e THOMAS, 1991). Houve, assim, importantes contribuições à *Large Deviation Theory*<sup>12</sup>.

Outra importante relação com Estatística surge a partir da condição de Máxima Entropia, terceiro axioma exposto na entropia de Shannon. Máxima Entropia pode ser entendido como um princípio para estimar, de modo menos incorreto possível, a distribuição de probabilidades de um processo com *prior data* conhecido. Pelo princípio, a melhor escolha da distribuição de probabilidades é aquela que maximiza a entropia do processo. Intuitivamente, o princípio proíbe aproximações sem evidência suficiente, já que a distribuição de maior entropia é a que faz menos suposições sobre a real distribuição dos dados. Esse mecanismo de escolha tem sido bastante utilizado em inferência Bayesiana, estimação de funções de densidade de probabilidade e outros.

Por parte da física, a análise do grau de desordem de um sistema é, de fato, a mensuração da quantidade de informação presente. Importantes aspectos desde o movimento browniano de partículas a dinâmica estelar foram bem caracterizados pela Teoria da Informação (GLEICK, 2011). Além disso, a mecânica quântica tem uma profunda relação com a informação, embora sua unidade de

---

<sup>12</sup> Teoria que formaliza a concentração de medidas assim como explora a noção da convergência de probabilidades mensuradas.



medida seja *qubits* e não os *bits* tradicionais. A possível corrida para o domínio da computação quântica (THE ECONOMIST, 2018) é um sinal da importância da integração dessas áreas.

Na ciência da computação, importantes resultados são intrinsicamente relacionados à TI. Caso emblemático, a Teoria Algorítmica da Informação foi concebida, por definição, através da combinação de ciência da computação e Teoria da Informação. Outra indicação da proximidade está no objeto de análise dessa ciência, o computador. Tal dispositivo é definido como uma máquina dedicada ao armazenamento, processamento e manipulação de informação. Seus componentes, em grande maioria, têm recursos cuja unidade de medida foi criada com base nos *bits*, o que mostra como a área evoluiu concomitante à formalização de Shannon.

Neurociência é uma das áreas cuja interação com Teoria da Informação mais prosperou. Termos como “sistema orgânico de processamento de informação” são muitas vezes utilizados em referência ao cérebro (GHAHRAMANI, 2006). Um dos campos correlatos, a psicologia cognitiva foi fundada através da interseção entre psicologia, ciência da computação e filosofia (GLEICK, 2011) e o ponto de interação é a informação. Frequentemente, estudos na área utilizam formulações de TI como método de pesquisa.

Um interessante uso de KL foi realizado em trabalho sobre rastreamento ocular. Itti e Baldi (2009) expuseram os participantes da pesquisa a diferentes vídeos e rastream o local de foco ocular. Ao mesmo tempo, os autores mensuraram a divergência de Kullback-Leiber entre os *pixels* da mesma região a cada momento no tempo e compararam com o foco dos participantes. Surpreendentemente, o dispositivo de rastreamento ocular mostrou que as pessoas tendem a visualizar regiões de maior KL. Assumindo um raciocínio Bayesiano, a pesquisa sugere que a atenção é focada em regiões que violam as expectativas subjetivas.

No campo da matemática, há importantes relações com Teoria da Informação. O Teorema de Brunn-Minkowski, que avalia desigualdades no volume

de subconjuntos compactos do espaço Euclidiano, possui relação próxima com desigualdades estudadas em TI (COVER e THOMAS, 1991). Uma delas, *Entropy Power Inequality* expõe uma relação de desigualdade entre a soma das entropias de duas variáveis aleatórias independentes e a entropia da soma dessas. Tal relação demonstra que variáveis aleatórias bem-comportadas são super-aditivas, resultado útil à análise de funções.

### 3.1. Economia e teoria da informação

Avaliar a influência de TI na economia pode conduzir a confusões. As duas áreas utilizam conceitos de mesmo nome, mas definidos de forma diferente. No campo originado por Shannon, informação é sinônimo de incerteza, cujo termo “incerteza” refere-se ao grau de desorganização ou aleatoriedade de um sistema físico (GLEICK, 2011). Em economia, incerteza e informação não são unicamente definidos.

Keynes (1936) e Knight (1921) propõem o uso do termo com fins de distingui-lo de risco. Incerteza seria o risco não probabilizável, no qual é impossível mensurar precisamente seu valor. A definição dos autores contradiz com o cálculo da quantidade informativa por meio de funções de probabilidade. Cabe lembrar que a TAI fornece medidas para a informação contida em objetos de probabilidade inacessível, sendo um possível modo de mensurar a incerteza knightiana sobre tomadas de decisões computáveis.

Ademais, como TI surgiu em estudos de comunicação e aplica-se a diversos outros campos, seria inusitado haver o termo “risco” para, digamos, a posição das moléculas de um líquido. Tais observações sugerem que o conceito de incerteza em TI e aquele relacionado à incerteza knightiana não são perfeitamente intercambiáveis.

Outra forma de definir a grandeza foi apresentada por Hirshleifer (1973), abordagem relacionada à Economia Informacional (*Information Economics*). Essa

definição interage melhor com TI, embora tenha ponto de vista simetricamente oposto à de Shannon. De modo pertinente, Hirshleifer relaciona incerteza com informação. Em sua proposta, incerteza mede a dispersão das probabilidades subjetivas sobre os “estados de mundo” enquanto informação seria o conjunto de eventos que poderia modificar tais probabilidades.

Seguindo o raciocínio de Hirshleifer, informação reduz incerteza, concentrando a distribuição de probabilidades em um conjunto de estados respaldados pela informação adquirida. Por outro lado, Shannon concebe informação como a própria medida de incerteza. Caso a informação, em TI, fosse aplicada ao contexto de Hirshleifer, essa seria a surpresa média ao observar que um determinado “estado de mundo” foi realizado (GARROUSTE, 2001). Na perspectiva de Teoria da Informação, há mais informação em uma distribuição de probabilidade dispersa do que em uma concentrada.

Um adendo: o conceito que melhor se adapta a descrição de Hirshleifer é a informação mútua. Ao conhecer uma outra variável correlacionada ao conjunto de “estados do mundo” a incerteza sobre o último efetivamente reduz. Nessa definição, a entropia condicional, sobre os “estados de mundo” dado a nova variável, é a quantidade de incerteza reduzida. Há aqui uma diferença teórica sutil, mas relevante. Ainda assim, o autor fez grande contribuição ao pensamento econômico e auxiliou na formação da Economia Informacional, maior responsável por tratar informação como uma variável central para a economia.

Economia Informacional analisa a influência da informação no comportamento econômico, evidenciando, entre outras questões, a fragilidade dos resultados tradicionais quando premissas mais realistas são adotadas, a exemplo da informação assimétrica entre agentes (STIGLITZ, 1985). Arrow (1996) expõe que informação é um bem econômico, no sentido que possui benefícios e custos. Suas características singulares, como a ausência de rivalidade e a enorme facilidade de transmissão, explicam os retornos crescentes de escala, que é um resultado semelhante ao de Romer (1986) sobre o papel das *ideias*.

O comportamento dos agentes, em *Information Economics*, é determinado pela função *payoff*, que mapeia as estratégias de ações para cada “estado de mundo” possível (ARROW, 1996). O conjunto de “estados de mundo” pode ser representado por uma variável aleatória, digamos  $S$ , no qual os agentes dificilmente têm acesso. Saber qual “estado de mundo” ocorre requer adquirir informação através de uma outra variável aleatória,  $Y$ , que detenha informação mútua.

Arrow (1986) revela que tanto o benefício quanto o custo da informação já foram modelados por meio de TI, embora a função custo seja mais comum. Nessas modelagens, o custo (fixo, dado que o bem só precisa ser adquirido uma vez) de  $Y$  é proporcional à informação mútua entre  $Y$  e  $S$ . Quanto maior a informação mútua,  $I(S; Y)$ , maior a precisão sobre o que ocorre em  $S$  e mais custoso é adquirir a variável, seguindo o princípio da escassez. O custo é independente da escala de renda do consumidor ou do tamanho da firma (ARROW, 1986).

Em relação ao benefício da informação, a análise considera que os elementos no espaço de possibilidades de  $Y$  produzem diferentes utilidades aos agentes. Saber qual  $y_i \in Y$  ocorreu pode definir a estratégia que o agente adota, interferindo diretamente em seu *payoff*. Ao contrário do custo, o benefício varia proporcionalmente com a renda. Segundo Arrow (1986), tal diferença entre benefício e custo é o motivo dos retornos crescentes à escala quando informação faz parte da modelagem econômica. Ainda, a facilidade de transmissão da informação impõe sérias dificuldades a provisão eficiente do bem, o que necessita de respostas incomuns aos mercados ditos competitivos.

Diante desse breve resumo sobre o campo, algumas considerações se fazem necessárias. Economia Informacional utiliza TI como intermediário para mensurar funções de interesse e não se apresenta como um subcampo da modelagem de Shannon. Especificamente, a análise sobre benefícios revela bem esse princípio. Teoria da Informação se preocupa com o comportamento sistêmico

de um processo e não expõe outra necessidade para os casos individuais a não ser o de calcular as medidas sistêmicas.

Outra evidência da relativa autonomia entre as áreas reside no próprio objetivo de *Informations Economics*, que é modelar a importância da informação no comportamento econômico. Assume-se, portanto, que os agentes interpretam a grandeza e produzam respostas em forma de ação. Nesse sentido, o campo aborda tanto o problema semântico quanto o problema da efetividade (Níveis B e C) presentes na comunicação, diferença crucial à abordagem de Shannon.

Expor algumas nuances envolvendo os termos incerteza e informação, presentes na economia, auxilia na boa compreensão das aplicações envolvendo TI. A esse respeito, houve contribuições robustas à Finanças. Cassetari (2003) utiliza o princípio de Máxima Entropia como medida de risco financeiro. Nessa modelagem, a solução de Markowitz pode ser encontrada assumindo que o risco possui comportamento gaussiano, o que aponta a Máxima Entropia como uma abordagem ainda mais ampla que a teoria de alocação de Markowitz convencional.

Ainda, é possível relacionar a taxa de crescimento da riqueza em um mercado financeiro com a entropia do mercado (COVER & THOMAS, 1991). Por esse motivo, Teoria da Informação permite definir estratégias de portfólio e, em particular, a estratégia ótima para mercados ergódicos, o que é, infelizmente, uma hipótese forte. Em suma, há uma teoria de portfólio formada no bojo da modelagem de Shannon.

Um proveitoso uso de TI se deu, também, na análise de desigualdade de renda. A literatura possui uma gama de indicadores que mensuram desigualdade, cada um com suas limitações e vantagens. Por esse motivo, Maasuomi (1993) recomenda uma abordagem axiomática para a escolha da medida, visto que definir propriedades de interesse aumenta a transparência sobre o motivo de seleção.

Após definir setes axiomas que expôs serem coerentes (simetria, normalização, princípio de transferências Pigou-Dalton, continuidade, homogeneidade, invariância de replicação e agregação consistente), Maasoumi (1993) revela que apenas um grupo de indicadores resta para seleção, os Índices de Entropia Generalizada. Esses índices são derivados de TI e representam a redundância presente em um sistema, no caso, a redundância sobre a renda.

Além dos trabalhos aplicados a fontes de dados convencionais, um novo uso começou a surgir, expandindo a compreensão de questões relevantes à literatura econômica. Essas aplicações se notabilizam pelo emprego de técnicas como *Topic Modelling* e *Machine Learning* a fim de garantir a formatação dos dados necessária para TI ser utilizável. A próxima seção nesse capítulo será destinada a apresentação de alguns destes artigos.

### **3.2. Aplicações recentes**

Barron et al. (2018) utilizam TI para rastrear a criação, transmissão e destruição de padrões em debates que ocorreram durante a Revolução Francesa, em específico, na Assembleia Nacional Constituinte (NCA). Esse método complementa análises histográficas previamente realizadas, mas oferece um mecanismo de avaliação quantitativo. Os autores analisaram mais de quarenta mil discursos e um pouco mais de mil oradores da NCA.

A partir do método de *Alocação latente de Dirichlet* (LDA), um aprendizado de máquina não supervisionado, modelou-se discursos como distribuições de probabilidade multinomiais a respeito de tópicos textuais. Cada tópico é gerado pela própria LDA, que avalia a semelhança estatística entre palavras (*tokens*) de um texto, ao passo que a quantidade de tópicos (exatamente cem) foi definida pelos autores. Após a modelagem, é possível analisar o grau de inovação e transitoriedade que cada padrão de discurso teve, utilizando-se KL. Entender a

definição de inovação e de transitoriedade, no contexto, necessita de uma pequena retomada com formalizações matemáticas.

Seja  $s_i^{(j)}$  a probabilidade do  $i$ -ésimo tópico representar o  $j$ -ésimo discurso. Ou seja,  $i$  é um indexador dos tópicos e  $j$  um indexador dos discursos. A divergência de Kullback-Leiber entre  $s^{(j)}$  e  $s^{(j-1)}$  é aqui expressa como:

$$KL(s^{(j-1)}) = \sum_{i=1}^K s_i^{(j)} \log_2 \frac{s_i^{(j)}}{s_i^{(j-1)}} \quad (3.1)$$

Onde  $K$  é a quantidade de tópicos. Note que a equação (3.1) apenas adiciona um indexador em relação a fórmula (2.14) e mantém a interpretação bayesiana de surpresa em  $s^{(j)}$  quando se espera  $s^{(j-1)}$ . A inovação de um discurso  $j$  com respeito aos  $w$  discursos anteriores,  $N_w(j)$ , representa o grau de surpresa média que  $j$  produziu sobre o contexto  $w$ :

$$N_w(j) = \frac{1}{w} \sum_{d=1}^w KL(s^{(j-d)}) \quad (3.2)$$

Quanto maior a inovação, maior a quebra das expectativas que  $j$  promove. Inovações altas no discurso  $j$  significam que o orador pode ter mudado de assunto ou mesmo tratado o problema discutido de modo significativamente diferente. A transitoriedade é o oposto de inovação, pois mede o grau de surpresa médio sobre os  $w$  eventos posteriores:

$$T_w(j) = \frac{1}{w} \sum_{d=1}^w KL(s^{(j+d)}) \quad (3.3)$$

Quanto maior a transitoriedade, mais rápido o padrão de  $j$  foi esquecido. Alta transitoriedade significa que os debatedores seguintes optaram por ignorar o padrão expresso em  $j$ , seja porque o debate acerca do assunto encerrou ou pela falta de propriedades idiossincráticas do orador (respaldo social, por exemplo).

Para medir o descompasso entre a inovação e a transitoriedade, criou-se também o indicador de ressonância:

$$R_w(j) = N_w(j) - T_w(j) \quad (3.4)$$

Ressonância mede a diferença entre inovação e transitoriedade. Quanto maior a ressonância, maior o efeito de  $j$  sobre o contexto  $w$ . Tal medida realça o papel tanto de oradores que possuíram alta inovação e baixa transitoriedade como de parlamentares cuja discursos detinham baixa inovação, mas transitoriedade menor ainda.

A mensuração da inovação nos padrões de discursos possui correlação com a concepção de Romer (1986) sobre inovação de ideias. É de se esperar que ideias diferentes necessitem de padrões diferentes para sua expressão. No entanto, o método utilizado enfatiza a avaliação de estratégias pelas quais oradores tentam influenciar o debate. Barron et al. (2018) expuseram conclusões instigantes a esse respeito.

Há uma forte relação positiva entre inovação e transitoriedade, de modo que discursos muito inovadores, em média, tendem a ser rapidamente esquecidos. A variância dos resultados, no entanto, expõe diferentes estratégias entre grupos políticos. Representantes políticos da esquerda, os jacobinos, possuem padrões de discursos mais inovadores, enquanto os conservadores foram responsáveis por manter o debate nos padrões já estabelecidos.

Alguns oradores se destacaram pelo efeito nos debates, quebrando o padrão esperado com alto valor de ressonância. Robespierre, jacobino famoso, produziu discursos de alta inovação e baixa transitoriedade, o que significa que o orador teve grande controle para determinar assuntos de debate. Representantes conservadores como Jean-Sifrein Maury e Jacques de Cazalès possuíram falas de baixa inovação, mas transitoriedade ainda menor, de modo que esses políticos conseguiam estabilizar o debate nos mesmos assuntos (Barron et al., 2018).

Além disso, foi possível analisar o impacto de cargos organizacionais recém-criados sobre o desenvolvimento de estratégias de discurso. Os presidentes da NCA marcaram as discussões pela alta transitoriedade de discursos. Isso porque tais integrantes recorrentemente resumiam os debates



realizados ao longo do dia, sem necessariamente participar da formulação dos argumentos.

Outra estrutura organizacional, os comitês de trabalho foram responsáveis por absorver a larga quantidade de demandas sem suprimir a deliberação de outros temas no plenário (Barron et al, 2018). Foi possível avaliar uma mudança importante no seu modo de operar, de modo que membros passaram a receber papel privilegiado nos debates. Por consequência, os discursos de membros ora encerravam debates ora determinavam o padrão dos tópicos dos próximos discursos. O trabalho dos autores, assim, auxiliou a entender a dinâmica dos debates construídos durante a Revolução Francesa, com potencial de replicação para os parlamentos atuais ou em outros conjuntos textuais.

Um uso bastante semelhante de Teoria da Informação ocorreu ao analisar as leituras que influenciaram Charles Darwin a criar sua Teoria da Evolução. Em função dos densos “jornais de leitura” que o cientista criou ao longo da vida, é possível avaliar as respostas de Darwin frente ao *trade-off* entre *exploration* e *exploitation* (Murdock et al, 2017). *Exploitation* refere-se ao aprofundamento de conhecimento em uma mesma área (processo de especialização) e *exploration* diz respeito a busca de diferentes formas de conhecimento (processo de descoberta). Esse *trade-off* está presente no desenvolvimento de novas tecnologias a partir da combinação de antigas (Youn et al., 2015).

Murdock et al (2017) utilizaram LDA para criar uma estrutura probabilística de tópicos sobre a base de livros, para depois aplicar a Divergência de Kullback-Leiver nestes. *Exploration*, aqui, foi proposta como os valores de KL acima da média, tanto entre dois textos próximos quanto entre um texto e os demais anteriores. Ao contrário, *exploitation* refere-se aos valores de KL abaixo da média, no mesmo contexto. A principal diferença técnica entre esse artigo e o anterior está no fato de que a análise das leituras de Darwin não avalia transitoriedade.

O artigo sustenta que Darwin variou de comportamento no decorrer de suas pesquisas. Com inicial estratégia de especialização em assuntos específicos,

*exploitation*, o biólogo passa a ter crescente tendência a variar tópicos de interesse, *exploration*. Esse resultado contradiz a estratégia de minimização de surpresa, que seria uma visão simplificada de como um agente navega para encontrar sua área de foco. Outro resultado notável na pesquisa, é que a ordem de leitura de Darwin é mais surpreendente (mais inovadora, seguindo (3.2)) que a ordem de publicação dos textos. Em suma, a pesquisa possui a limitação inerente de estudar um único indivíduo, mas estimula a análise de outros cientistas ou artistas para comparação.

Por fim, um trabalho que estende o estudo de combinação entre estratégias foi desenvolvido por Mueller (2019). O pesquisador analisa a influência musical presente em compositores de sucesso. Cada compositor anunciou suas principais referências em entrevista ao *Song Explorer Podcast*. Caracterizadas tais relações, Mueller (2019) estima, por meio de medida derivada de TI, o grau de proximidade entre as músicas produzidas pelo compositor e suas influências.

Essa proximidade é então utilizada para explicar o sucesso dos músicos em termos de visualizações em plataforma de vídeos, o *Youtube*. O emblemático resultado confirma que o impacto musical é otimizado por uma estratégia mista entre a convencionalidade de cada compositor, em seu gênero de origem, e a inovação presente ao absorver padrões de gêneros distantes do seu inicial.

## **4. Aplicação Empírica**

### **4.1. O Encontro Nacional de Economia**

As aplicações recentes envolvendo TI, ainda que escassas, revelam como o uso de dados não convencionais podem promover resultados poderosos. Para ampliar essa literatura, utilizou-se da formalização criada em Barron et al. (2018) a fim de analisar o grau de inovação, transitoriedade e ressonância da pesquisa econômica. Em específico, tais medidas foram aplicadas a um *corpus* linguístico de artigos apresentados em evento promovido pela ANPEC.

O Encontro Nacional de Economia é reconhecido como um dos mais importantes congressos de Economia do país. Realizado anualmente, o Encontro seleciona trabalhos que exploram as fronteiras da pesquisa econômica para que os autores apresentem seus achados em meio a comunidade de pesquisadores e

profissionais. Esses artigos são posteriormente divulgados nos Anais do Encontro e disponibilizados no site oficial da instituição<sup>13</sup>.

A partir de 2013, a ANPEC reorganizou os trabalhos admitidos em 13 áreas:

- Área 1: História do pensamento econômico e metodologia;
- Área 2: Economia política;
- Área 3: História econômica;
- Área 4: Macroeconomia, economia monetária e finanças;
- Área 5: Economia do setor público;
- Área 6: Crescimento, desenvolvimento econômico e instituições;
- Área 7: Economia internacional;
- Área 8: Microeconomia, métodos quantitativos e finanças;
- Área 9: Economia industrial e da tecnologia;
- Área 10: Economia regional e urbana;
- Área 11: Economia agrícola e do meio ambiente;
- Área 12: Economia social e demografia econômica;
- Área 13: Economia do trabalho.

O presente trabalho utilizou 1676 *papers* publicados entre 2013 e 2019 nos Anais do Encontro. Aplicou-se o método de modelagem de tópicos, LDA, nesse *corpus* com o intuito de mensurar as variáveis de interesse. A partir da classificação de áreas adotada pela ANPEC, foi possível determinar quais foram as variações em termos de inovação, transitoriedade e ressonância entre artigos da mesma área. Além disso, verificou-se um comportamento inercial na relação entre os artigos ao longo do tempo.

## 4.2. Escopo da aplicação

---

<sup>13</sup> <https://en.anpec.org.br/previous-editions.php>

A análise foi implementada no *Python*, uma das melhores linguagens de programação para Processamento Natural de Linguagem. Inicialmente, as *urls* contendo os *hyperlinks* de *download*, nome dos autores e área do artigo foram raspadas e os resultados salvos como uma base de dados estruturados<sup>14</sup>. Do total de 1681 artigos publicados entre 2013 e 2019, três destes possuíam problemas para *download* posto que os arquivos estavam danificados<sup>15</sup>.

Os documentos baixados foram convertidos para arquivo de texto. Posteriormente, executou-se um algoritmo de detecção da linguagem textual. Identificar a língua do texto é fundamental uma vez que as bibliotecas convencionais de LDA foram produzidos para cenários de *corpus* unilíngues. Entre os 1678 artigos restantes, dois destes foram escritos em espanhol, o que se configura quantidade excessivamente pequena de textos para realização de modelagem de tópicos. Por esse motivo, esses artigos foram retirados da base, restando 1100 artigos em português e 576 em inglês. O *corpus* foi dividido pela língua em dois *corpora* e os procedimentos foram realizados separadamente em cada conjunto.

A próxima etapa da implementação consistiu no pré-processamento dos textos, determinante fundamental da qualidade da modelagem. Retirou-se as referências bibliográficas, caracteres mal codificados<sup>16</sup>, caracteres especiais, dígitos, e-mails, referências a sites, pontuações (com exceção dos acentos) e palavras com menos do que três letras. As palavras restantes foram convertidas para minúsculas. Foi criado, também, uma lista de *stopwords*<sup>17</sup> com cerca de 1900

---

<sup>14</sup> No ano de 2015, a ANPEC não disponibilizou a organização dos arquivos em áreas nem o nome dos autores. No entanto, o *hyperlink* para *download* do artigo contém uma sigla, que se refere a área do artigo.

Exemplo: [https://www.anpec.org.br/encontro/2015/submissao/files\\_l/i7db657db014e3a564f5a9573d3bd37137.pdf](https://www.anpec.org.br/encontro/2015/submissao/files_l/i7db657db014e3a564f5a9573d3bd37137.pdf)

O termo “i7” após “files\_l/” se refere à sétima área definida pela ANPEC. Assim, a base conteve área de todos os artigos, mas faltou o nome dos autores para documentos de 2015.

<sup>15</sup> Is tracking beneficial? Study of tracking using peer effects (2016); Efeitos do investimento em capital intangível e patentes no valor das empresas brasileiras (2013); The ins and outs of unemployment in a dual labor market (2016).

<sup>16</sup> Trata-se de caracteres cuja codificação em *UTF-8* sofreu problemas. Foi um problema pouco frequente (cerca de 4 documentos), mas dificultaria a modelagem de tópicos dos artigos.

<sup>17</sup> *Stopwords* são palavras que devem ser retiradas para implementação de algoritmos de modelagem de tópicos. Entende-se que tais palavras atrapalham na modelagem dos tópicos

palavras. O pré-processamento seguiu as boas práticas expostas em Asmussen e Møller (2019).

Para auxiliar na compreensão dos padrões particulares de produção acadêmica, foram criadas *Wordclouds* contendo as palavras mais frequentes de cada área (Figuras 2 e 3). Essa ferramenta permite avaliar a qualidade do pré-processamento dos *corpora* e serve de referência para análises envolvendo o comportamento dessas áreas.

É possível perceber o grau de semelhança entre a Área 5 (Economia do setor público) e a Área 10 (Economia regional e urbana), embora cada área enfatize perguntas de pesquisa diferentes. A Área 1 (História do pensamento econômico e metodologia) aparenta possuir o escopo mais amplo entre todas. Suas palavras mais frequentes permitem a caracterização de um grupo diverso de assuntos para pesquisa. Esse fato pode significar um padrão mais inovador nessa área, já que há uma maior probabilidade de artigos apresentarem padrões de escrita diferentes.

---

(exemplos: artigos, pronomes e palavras que possuem com alta frequência em todos os documentos de um *corpus*).



Figura 3 – Wordcloud dos artigos publicados em inglês, por área

Área 1 - História do Pensamento Econômico e Metodologia



(2013-2019) Área 2 - Economia Política



Área 3 - História Econômica



Área 4 - Macroeconomia, Economia Monetária e Finanças



Área 5 - Economia do Setor Público



Área 6 - Crescimento, Desenvolvimento Econômico e Instituições



Área 7 - Economia Internacional



Área 8 - Microeconomia, Métodos Quantitativos e Finanças



Área 9 - Economia Industrial e da Tecnologia



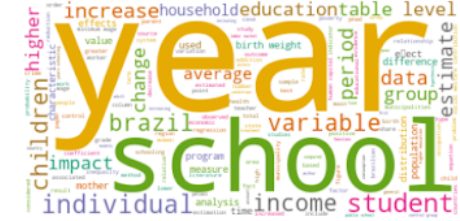
Área 10 - Economia Regional e Urbana



Área 11 - Economia Agrícola e do Meio Ambiente



Área 12 - Economia Social e Demografia Econômica



Área 13 - Economia do Trabalho



Fonte: Autor



Antes de aplicar o algoritmo de Alocação Latente de Dirichlet, os *corpora* linguísticos devem passar por um processo de vetorização. Optou-se pelo algoritmo de vetorização baseado em *Term Frequency–Inverse Document Frequency* (TF-IDF), concomitante a um processo de Stemização (*stemming*). A vetorização utilizou as 12.000 palavras mais comuns de cada *corpus* (separadamente) e retirou-se as palavras que ocorriam em mais de 90% dos artigos.

TF-IDF é uma medida estatística que pondera a frequência relativa de palavras em um documento com sua frequência relativa no *corpus* como um todo. Seu objetivo consiste em reduzir o peso de palavras que existem em muitos documentos e aumentar o peso de termos que aparecem particularmente com frequência em um determinado documento. Intuitivamente, a medida restringe a importância de palavras como “análise” (aparecem na maioria dos artigos) e amplia o papel de palavras como “desmatamento” (aparecem em um número limitado) na modelagem de tópicos.

Stemização é o processo de remoção de flexões linguísticas para identificar a raiz de uma palavra. Tal procedimento é importante para que palavras no singular e no plural, ou mesmo flexões verbais, sejam reconhecidas como originárias do mesmo termo. Atuais programas de Stemização para a língua portuguesa ainda necessitam de bastante ajuste para termos irregulares, não obstante o programa<sup>18</sup> utilizado é reconhecido como um dos mais precisos.

Após a vetorização, a modelagem de tópicos por LDA foi finalmente executada. Utilizou-se a biblioteca *gensim*, que fornece um método computacionalmente eficiente para otimização das palavras nos tópicos, o *Online Variational Bayes* (OVB). Baseado em otimização estocástica, OVB converge mais rapidamente a um equilíbrio comparado a outros versões de cálculo bayesiano (Hoffman et al., 2010).

Uma questão bastante difícil é a definição da quantidade de tópicos. Pelos critérios de seleção convencionais em aplicações do gênero, *Perplexity* e *Topic Coherence Score* (ambos derivados de TI), maior quantidade de tópicos tende a significar melhores modelos. Tanto no *corpus* de inglês quanto no de português, entre

---

<sup>18</sup> <https://snowballstem.org/>. Essa é a biblioteca *PyStemmer* no *Python*

10, 13, 15, 20, 30, 50 e 100 tópicos, o critério *Perplexity* favorece modelos com a maior quantidade destes, ao passo que o *Topic Coherence Score* apresenta diferenças marginais entre 30 e 100 tópicos, favorecendo o primeiro modelo no *corpus* em português.

*Topic Modelling* ainda é uma literatura muito recente e falta resultados robustos sobre problemas como a otimização da quantidade de tópicos. Optou-se pela modelagem contendo 30 tópicos para ambos os *corpora*, embora o argumento para outras escolhas seja igualmente válido<sup>19</sup>.

Por fim, com a distribuição de probabilidades sobre tópicos criada, aplicou-se as fórmulas de inovação –  $N$  (3.2), de transitoriedade -  $T$  (3.3) e de ressonância -  $R$  (3.4) em relação a artigos de mesma área e de mesmo *corpus*. O parâmetro  $w$ , que indica a escala do cálculo, foi definido como o valor máximo, ou seja, engloba todos os artigos anteriores, para o cálculo da inovação, e todos os artigos posteriores, para o cálculo da transitoriedade. A Ressonância representa a diferença entre esses valores.

Uma limitação da base de dados utilizada reside no fato de que não é possível saber a ordem de criação dos artigos publicados no mesmo ano. A fim de contornar a limitação, foi criado uma função que ordena aleatoriamente os artigos de mesma data e simula os resultados de (3.2), (3.3) e (3.4). As operações são realizadas dezenas de vezes e utiliza-se a mediana de todas as simulações como o valor final. Essa solução é bastante parecida com aquela apresentada por Murdock et al (2017) para lidar com o mesmo problema. Após todos os cálculos os *corpora* voltam a ser integrados em uma mesma base de dados.

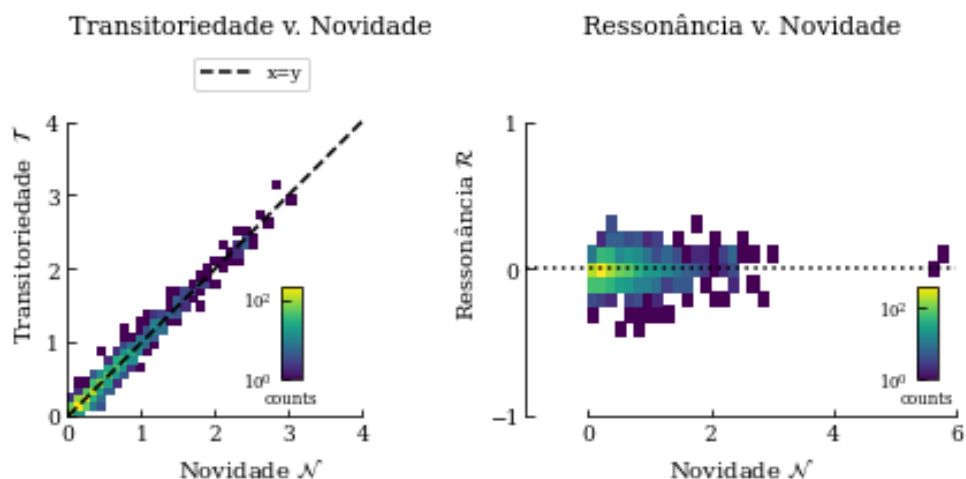
---

<sup>19</sup> Um ponto importante é que as principais conclusões deste trabalho se mantiveram consistentes com modelagem de 30 e de 100 tópicos. Houve algumas variações em relação à classificação das áreas em ordem de inovação, embora as duas áreas mais inovadoras e as duas menos inovadoras se mantiveram as mesmas.

### 4.3. Resultados e discussões

O resultado evidencia ampla variedade de estratégias de inovação. A figura 4 resume o comportamento do sistema, revelando a alta correlação entre inovação e transitoriedade. É possível perceber que a maior parte dos documentos se situam próximos ao eixo de simetria (primeiro gráfico da Figura 4), de modo que publicações inovadoras são punidas com o aumento, recorrentemente, equivalente de sua transitoriedade. Observando a relação entre ressonância e inovação, fica frágil a sustentação de algum viés sistêmico em favor de pesquisas inovadoras.

Figura 4 – Gráficos de densidade das variáveis de interesse



Fonte: Autor, com códigos de Barron et al. (2018)

Há um comportamento inercial na produção acadêmica, no qual cada artigo tem pouco poder para influenciar os padrões de pesquisas de seus sucessores. Nesse sentido, o processo de escolha, por área, dos trabalhos admitidos no Encontro pode ser caracterizado como uma seleção de documentos em que o conteúdo geral é relativamente próximo entre si, ao passo que cada artigo, individualmente, possui algum grau de independência com seus antecessores/sucessores. Quando algum destes

trabalhos desenvolve padrões mais inovadores que a média, os próximos artigos selecionados tendem a rapidamente retomar ao equilíbrio original.

Ainda assim, percebe-se um comportamento médio distinto entre as áreas do conhecimento. Conforme é possível visualizar pela Tabela 1, o campo de “Microeconomia, métodos quantitativos e finanças” possui a maior média de inovação enquanto “Economia Política” fica em último lugar<sup>20</sup>. A figura 5 expõe uma comparação entre as áreas em termos de inovação e transitoriedade. É possível perceber que nenhuma área se insere nos quadrantes de baixa transitoriedade e alta novidade ou no de alta transitoriedade e baixa novidade.<sup>21</sup> Chama a atenção, também, a primeira área, “História do pensamento econômico e metodologia”, que reproduziu a hipótese prevista com a visualização dos *Wordsclouds*. Em função de sua grande variedade de assuntos de análise, a área possui elevado grau de inovação e de transitoriedade.

**Tabela 1 – Média e desvio padrão (DP) de , por área**

Áreas do conhecimento			T		R	
	Médi a	DP	Média	DP	Média	DP
Área 1 - História do pensamento econômico e metodologia	0,691 4	0,4565	<b>0,7144</b>	0,4490	-0,0237	0,056
Área 2 - Economia política	<b>0,233</b> 2	0,2569	0,2502	0,2568	-0,0176	0,028
Área 3 - História econômica	0,372 6	0,2510	0,3809	0,2649	-0,0046	0,076
Área 4 - Macroeconomia, economia monetária e finanças	0,592 8	0,7124	0,5603	0,7199	<b>0,0331</b>	0,101
Área 5 - Economia do setor público	0,557 7	0,5247	0,5941	0,5239	<b>-0,0372</b>	0,077
Área 6 - Crescimento, desenvolvimento econômico e instituições	0,253 8	0,2406	<b>0,2400</b>	0,2490	0,0141	0,035
Área 7 - Economia internacional	0,281 3	0,3419	0,3118	0,3417	-0,0313	0,045
Área 8 - Microeconomia, métodos quantitativos e finanças	<b>0,701</b> 1	0,4367	0,6943	0,4250	0,0077	0,079
Área 9 - Economia industrial e da tecnologia	0,336 4	0,3861	0,3529	0,3664	-0,0170	0,047
Área 10 - Economia regional e urbana	0,495 6	0,3635	0,5117	0,3677	-0,0165	0,055
Área 11 - Economia agrícola e do meio ambiente	0,546 1	0,3873	0,5703	0,4107	-0,0247	0,065

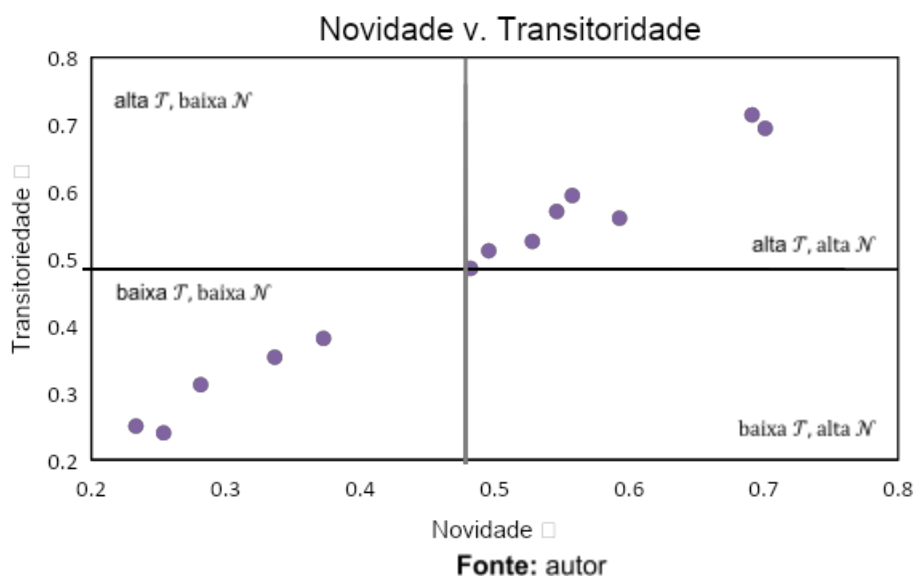
<sup>20</sup> Importante salientar que a medida de inovação utilizada não possui relação com a qualidade das pesquisas. De fato, inovação, nesse caso, é um sinônimo de diversidade padrões de escrita, ou seja, mensura com que grau artigos mudam a importância relativa de cada palavra.

<sup>21</sup> As linhas de fronteiras foram determinadas a partir da média geral dos artigos, ou seja, N = 0,4783 e T = 0,4841.

Área 12 - Economia social e demografia econômica	0,528 0	0,4646	0,5256	0,4503	0,0025	0,039
Área 13 - Economia do trabalho	0,482 1	0,4547	0,4854	0,4367	-0,0029	0,089
Geral	0,478 3	0,4698	0,4841	0,4677	-0,0056	0,068

Fonte: Autor

Figura 5 – Comparação entre as áreas



Os dados apresentados na Tabela 1 ratificam a interpretação de que a inovação é rapidamente dissipada, uma vez que não há valor de ressonância, em nenhuma área, estatisticamente superior à média geral<sup>22</sup>. O comportamento das áreas, em termos de inovação e transitoriedade, necessita de outras pesquisas para futura validação, mas trata-se de um achado robusto.

Uma última análise ocorreu para os artigos que se situam nas caudas da inovação. Os cinco artigos de menor e de maior inovação foram expostos na Tabela 2, apontando alguns padrões importantes. Em primeiro lugar, os dois artigos mais inovadores da base possuem valores surpreendentemente altos. Uma possível

<sup>22</sup> A área com maior valor médio de ressonância, “Macroeconomia, economia monetária e finanças, possui escore padrão de  $Z = (0,331 - (-0,0056))/0,0688 \cong 0,5625$ , valor muito pequeno.

explicação para o fenômeno é que ambos os artigos podem ter estruturas textuais mais semelhantes a áreas diferentes daquela em que foram admitidos.

Por exemplo, o trabalho “Regressão Quantílica e Var: uma aplicação de quantis condicionais extremos para os retornos relativos ao Ibovespa e Petrobrás” (Câmpelo et al., 2017) possui muitos termos associados à microeconomia (retorno, risco, Ibovespa) enquanto “*Does mixed frequency vector error correction model add relevant information to exchange misalignment calculus? Evidence for United States*” (Marçal et al., 2014) faz uso recorrente de termos empregados em Economia Internacional. Assim, o LDA pode ter gerado distribuições de probabilidade bem distintos dos demais artigos admitidos na Área 4, o que resultou em alto valor de inovação.

Os artigos menos inovadores foram todos da mesma área, Economia Política. Apesar de ser um instrumental em constante progresso, é possível que os termos comuns a essa literatura sejam particularmente raros nos demais campos. Tal fator dificultaria a captação da diversidade do campo em modelagens de tópicos.

**Tabela 2 – Caudas da Inovação (5 artigos)**

<b>Títulos</b>	<b>Autores</b>	<b>Área</b>			
<b>Artigos de menor inovação</b>					
As possíveis linhas mestras do debate marxista sobre a grande recessão (2007-2009)	Alex Wilhans Antonio Palludeto	Área 2: Economia política	0,0324	0,0533	-0,0209
Distribuição e crescimento em uma economia fechada e com alta inflação: o caso brasileiro do Pós II Guerra	Carmem Feijó	Área 2: Economia política	0,0329	0,0560	-0,0241
Da posição e da deposição histórica do valor	Eleuterio Fernando Da Silva Prado	Área 2: Economia política	0,0334	0,0544	-0,0245
Desmaterialização desmistificada: o potencial poupador da tecnologia e suas implicações ambientais	Eduardo Sá Barreto	Área 2: Economia política	0,0336	0,0534	-0,0240
Notas sobre a crise de 2008: transição para uma nova fase do capitalismo?	Eduardo Da Motta e Albuquerque	Área 2: Economia política	0,0389	0,0571	-0,0216
<b>Artigos de maior inovação</b>					
Regressão Quantílica e Var: uma aplicação de quantis condicionais extremos para os retornos relativos ao Ibovespa e Petrobrás	Ana Katarina T. De Novaes Campelo; Wilton Bernardino da Silva; Renan O. Regis; Rafael Moura Azevedo	Área 4: Macroeconomia, Economia Monetária e Finanças	5,7138	5,5842	0,1277
Does mixed frequency vector error correction model add relevant information to exchange misalignment calculus? Evidence for United States	Diogo De Prince Mendonça	Área 4: Macroeconomia, Economia Monetária e Finanças	5,5724	5,5385	0,0364
Relação entre exposição à violência e habilidades socioemocionais: O caso dos estudantes de Sertãozinho (Sp)	Wander Plassa; Carolina Moraes Sarmento; Luiz Guilherme Scorzafave; Daniel D. Dos Santos	Área 12: Economia Social e Demografia Econômica	2,9995	2,9337	0,0687
Outsiders na política melhoram a gestão municipal?	Henrique Augusto C. Fernandez Hott; Naercio Aquino Menezes Filho	Área 5: Economia do Setor Público	2,8821	3,1057	-0,2367
Yardstick Competition e a disciplina eleitoral no Programa Bolsa Família	Mario Rubem do Coutto Bastos; Bernardo P. Machado Mueller	Área 5: Economia do setor público	2,7229	2,6687	0,0543

Fonte: Autor

## 5. Conclusões

Teoria da Informação relaciona a quantidade informativa de um sistema com o grau de incerteza ou aleatoriedade presente nesse. Tal modelagem surgiu como solução de um problema presente em sistemas comunicativos, expandindo profundamente a compreensão dos fenômenos que envolvem a comunicação. Devido as características universalizantes dessa modelagem, TI foi aplicada em diversos saberes científicos, a exemplo da Economia.

Na análise econômica, informação e incerteza não são conceitos unicamente definidos, o que pode levar a confusões. Especificamente, em *Information Economics*, há distinções relevantes. A definição sustentada em Economia Informacional aborda o valor semântico presente na informação, o que é fora do escopo de TI.

Não obstante, o presente trabalho tentou mostrar a enorme capacidade que Teoria da Informação possui para auxiliar o desenvolvimento científico da Economia. Foi exposto algumas aplicações em diferentes áreas como Finanças, Desigualdade Econômica, Econometria e Inovação, que obtiveram resultados estimulantes para novas pesquisas.

Em específico, o artigo realizou uma aplicação singular através do uso de um conceito derivado da entropia de Shannon, a Divergência de Kullback-Leiber (KLD). Com essa modelagem, foi possível avaliar o grau de inovação, transitoriedade e ressonância nos artigos disponíveis pelos Anais dos Encontros Nacionais de Economia (ANPEC). Verificou-se um poderoso padrão de dissipação da inovação, em que artigos inovadores eram igualmente transitórios. O método empregou uma modelagem de tópicos a partir da Alocação Latente de Dirichlet (LDA), conhecido método de aprendizado de máquinas não supervisionado.



## Referências

AFTAB O.; CHEUNG P.; KIM A., THAKKAR S. YEDDANAPUDI N. Information Theory and the Digital Age. 6.933—Final Paper, The Structure of Engineering Revolutions, Massachusetts Institute of Technology, Cambridge, 2001.

ARROW K. J. The Economics of Information: An Exposition. *Empirica* 23: 119-28,1996.

ASMUSSEN C.B., MØLLER C. Smart literature review: a practical topic modelling approach to exploratory literature review. *J. Big Data*. 2019;6(1)

BARON A. T. J.; HUANG, J.; SPANG, R. L.; DEDEO Simon. Individuals, institutions and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*. 115 (18), 2018.

CAMPÊLO Ana K.; SILVA Wilton B.; Silva Renan O.; AZEVEDO R. M. Regressão Quantílica e VaR: Uma Aplicação de Quantis Condicionais Extremos para os Retornos Relativos ao IBOVESPA e Petrobrás. 45° Encontro Nacional de Economia - Trabalhos selecionados. 2015.

CASSETTARI, A. O princípio da máxima entropia e a moderna teoria das carteiras. *Revista Brasileira de Finanças*, 1(2), 271-300, 2003.

CHURCH, A. An Unsolvable Problem of Elementary Number Theory. *Amer. J. Math.*58, 345-363, 1936.

COVER T.M. THOMAS J.A. *Elements of information theory*. John Wiley, 1991.

DEDEO, Simon. *Information theory for intelligent people*, 2016. <http://tuvalu.santafe.edu/~simon/it.pdf>.

DRETSKE, F. *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press, 1981.

EVREN, A TUNA, E. On some properties based on goodness of fit measures based on statistical entropy." International Journal of Research & Reviews in Applied Sciences. Vol. 13 (1), 192-205, 2012.

FLORIDI, L. "What Is the Philosophy of Information?" *Metaphilosophy*, 33(1–2): 123–145, 2002.

GARROUSTE, Pierre "What Economics Borrows from the Statistical Theory of Information?" In: Petit Pascal (eds) *Economics and Information*. Springer, Boston, MA, 2001.

GHAHRAMANI, Z. Information Theory. *Encyclopedia of Cognitive Science*, Ed. L. Nadel, 2006.

GLEICK, J. *The information: a history, a theory, a flood*. Fourth Estate, London, UK, 2011.

GRÜNWARD, Peter D. VITÁNYI M. B. Paul. Algorithmic Information Theory, in Adriaans and van Benthem 2008b: 281–317, 2008. doi:10.1016/B978-0-444-51726-5.50013-3.

HIRSHLEIFER J. Where are we in the theory of information? *American Economic Review*, p. 31-39, May 1973.

HOFFMAN M., BLEI David M., and BACH Francis. Online learning for latent Dirichlet allocation. In *Advances in Neural Information Processing Systems 24*, 2010.

ITTI, Laurent BALDI Pierre. Bayesian surprise attracts human attention. *Vision Research*, 49(10): 1295-1306, 2009.

KEYNES, J. M. *The General Theory of Employment, Interest and Money*. London: Macmillan, 1936.

KNIGHT, F. H. *Risk, uncertainty and profit*. New York: Hart, Schaffner and Marx, 1921.

KOLMOGOROV, A. Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics*, 2, 157–168, 1968. doi:10.1080/0020716680880303.

LEVIN A. Leonid. Some Theorems on the Algorithmic Approach to Probability Theory and Information Theory, Dissertation in Mathematics, Moscow, 1971.

MAASUOMI, E. A compendium to information theory in economics and econometrics. *Econometric Reviews*, 137-181, 1993.

MANNING, C. SCHUTZE, H. Foundations of Statistical Natural Language Processing. The MIT Press, Cambridge, US, 1999.

MÁQUINA DE TURING UNIVERSAL. Em: Wikipédia: a enciclopédia livre. Disponível em: [https://pt.wikipedia.org/wiki/M%C3%A1quina\\_de\\_Turing\\_universal](https://pt.wikipedia.org/wiki/M%C3%A1quina_de_Turing_universal). Acesso em: 20 fev. 2020.

MARTIGNON, L. Information Theory. In: *International Encyclopedia of the Social Behavioral Sciences*, Ed. Neil J. Smelser, Paul B. Baltes, 7476-7480. Oxford: Pergamon, 2001.

MARÇAL E. M, ZIMMERMANN Beatrice, PRINCE Diogo, MERLIN G. Does mixed frequency vector error correction model add relevant information to exchange misalignment calculus? Evidence for United States. 42º Encontro Nacional de Economia - Trabalhos selecionados. 2014.

MURDOCK J., ALLEN C., DEDEO S. Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition* 159:117–126, 2017.

MUELLER, Bernardo. Where 'd You Get that Idea? Determinants of Creativity and Impact in Popular Music. Working paper, Universidade de Brasília, 2019.

RHOADES, S. The Herfindahl-Hirschman Index. *Federal Reserve Bulletin* 79(3): 188-189, 1993.

ROMER Paul M. Increasing Returns and Long-Run Growth. *Journal of Political Economy* 94: 1002—1037, 1986.

SHANNON, C. WEAVER, W. The Mathematical Theory of Communication, Urbana, IL: University of Illinois Press, 1949.

SHANNON, Claude E. A Mathematical Theory of Communication. Bell System Technical Journal. 27 (3): 379–423, July 1948.

SHANNON, Claude E. The bandwagon. IRE Transactions on Information Theory, 2(1):3-3, March 1956.

STIGLITZ, J. E. Information and economic analysis: a perspective. The Economic Journal Supplement, vol. 95, 1985.

THE RACE IS ON TO DOMINATE QUANTUM COMPUTING. The Economist, Business, 18 de agosto de 2018. Disponível em: <https://www.economist.com/business/2018/08/18/the-race-is-on-to-dominate-quantum-computing>. Acesso em: 02 dez. 2019.

THURNER Stefan, COROMINAS-MURTRA Bernat, HANEL Rudolf. Three faces of entropy for complex systems: Information, thermodynamics, and the maximum entropy principle. Phys. Rev. E, 96:032124, 2017.

TRIBUS M. MCIRVINE E.C. Energy and information, Scientific American, 224,178–184, 1971.

WHEELER J. Archibald. At Home in Universe. American Institute of Physics Press, New York, 1994.

XIE, L. L, KUMAR P. R. A network information theory for wireless communication: Scaling laws and optimal operation. IEEE Trans. Inf. Theory, vol. 50, no. 5, pp.748–767, May 2004.

YOUN, H., STRUMSKY, D., BETTENCOURT, L. M., LOBO, J. Invention as a combinatorial process: Evidence from US Patents. Journal of the Royal Society, 12, 1-8, 2015. <http://dx.doi.org/10.1098/rsif.2015.0272>.