

Universidade Estadual de Maringá

Departamento de Física

Alvaro Franco Martins

Dinâmica de Redes de Corrupção Política

Maringá, 25 de março de 2021.

Universidade Estadual de Maringá

Departamento de Física

Alvaro Franco Martins

Dinâmica de Redes de Corrupção Política

Dissertação apresentada ao Programa de Pós-Graduação em Física da Universidade Estadual de Maringá como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro

Maringá, 25 de março de 2021.

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

M386d

Martins, Alvaro Franco

Dinâmica de redes de corrupção política / Alvaro Franco Martins. -- Maringá, PR, 2021.
75 f.: il. color., figs., tabs.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro.

Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Ciências
Exatas, Departamento de Física, Programa de Pós-Graduação em Física, 2021.

1. Corrupção política - Espanha. 2. Redes criminosas. 3. Ciência de redes. 4. Sistemas
complexos - Física estatística. I. Ribeiro, Haroldo Valentin, orient. II. Universidade Estadual
de Maringá. Centro de Ciências Exatas. Departamento de Física. Programa de Pós-
Graduação em Física. III. Título.

CDD 23.ed. 364.1323

Síntese Raquel Eleutério - CRB 9/1641

Agradecimentos

Reconheço os esforços de todos que me apoiaram durante o desenvolvimento desta dissertação. Meu muito obrigado à minha família e aos meus amigos. Em especial, agradeço ao Prof. Dr. Haroldo Valentin Ribeiro pela orientação neste trabalho.

Resumo

A ciência de redes representa uma abordagem de sistemas complexos capaz de estudar fenômenos sociais de forma estatística por meio de elementos simples. Neste trabalho, estudamos uma rede de corrupção política da Espanha e comparamos nossos resultados com um estudo existente sobre uma rede de corrupção política do Brasil. Nesse contexto, os implicados em escândalos de corrupção são tratados como vértices e dois indivíduos estão conectados se estiveram envolvidos em um mesmo escândalo. Para realizar nossas investigações, contamos com uma base de dados extraída do *site casos-aislados*, uma iniciativa sem fins lucrativos e não vinculada a partidos políticos que apresenta um compilado atualizado de informações sobre escândalos de corrupção espanhóis. Em uma primeira análise, usamos essas informações para explorar aspectos demográficos da corrupção espanhola. Em seguida, criamos a rede de corrupção e estudamos algumas de suas principais características estruturais por meio de medidas de centralidade e ataques direcionados a vértices específicos. Dentre esses resultados, destacam-se o papel desempenhado por envolvidos reincidientes em práticas ilícitas e a existência de uma polarização partidária na rede. Nossos achados indicam que essa polarização pode ter influenciado o comportamento de diversos aspectos observados na rede. Além disso, com a finalidade de entender os processos responsáveis pela evolução dessa rede, levamos em conta seu aspecto temporal e encontramos um fenômeno tipo coalescência responsável pelo estado final da rede de corrupção espanhola. Notamos que esse processo resultou da junção das partes politicamente polarizadas da rede, causando um aumento abrupto no tamanho da maior componente. Em uma penúltima análise, usamos um modelo conhecido como *stochastic block modeling* para extrair as comunidades (ou módulos) da rede. Com essa informação, comparamos o número de escândalos por módulo entre as redes de corrupção brasileira e espanhola. Por fim, motivados por nossos achados empíricos e pelo papel de envolvidos reincidientes em atividades ilícitas, propomos um modelo computacional para a formação e evolução de redes de corrupção política. De forma surpreendente, mesmo levando em conta ingredientes simples, esse modelo foi capaz de replicar vários de nossos achados empíricos e promover um entendimento mais completo sobre o comportamento de redes de corrupção.

Palavras-chave: Corrupção. Espanha. Redes de corrupção. Redes Criminosas. Ciência de Redes. Sistemas Complexos. Ciência de Dados.

Abstract

Network science represents a complex systems approach that enables statistical analysis of social phenomena by using simple elements. In this work, we study a Spanish political corruption network and compare our findings with a previous study about a Brazilian political corruption network. In this context, people involved in corruption scandals are considered as vertices, and two individuals are connected if they got involved in the same scandal. To do so, we have obtained a dataset from *casos-aislados*, a non-profit website that is not linked to any political parties and contains updated information on Spanish corruption scandals. In a first analysis, we have used this information to explore demographic aspects of Spanish corruption. We then have built a corruption network and studied some of its main structural features through measures of centrality and attacks directed at specific nodes. Among our results, we highlight the repeat offenders' role and the political party polarization of the network. Our findings indicate that this polarization may have affected the behavior of several aspects reported throughout our study. To understand the processes responsible for the evolution of this network, we have considered its time-dependent aspect and identified a coalescence-like phenomenon responsible for the final stage of the Spanish corruption network. We have noticed that this process results from joining the politically polarized parts of the network, promoting an abrupt increase in its largest component. In another analysis, we apply the stochastic block modeling approach to extract the modular structure of the Spanish corruption network. Using this information, we have compared the number of scandals per module between both Brazilian and Spanish corruption networks. Lastly, motivated by our empirical findings and the role of repeat offenders in illegal activities, we have proposed a computational model for the formation and evolution of political corruption networks. Although our model accounts only for simple ingredients, we show that it can replicate several of our empirical findings and promote a more in-depth understanding of the behavior of corruption networks.

Keywords: Corruption. Spain. Corruption Networks. Crime Networks. Network Science. Complex Systems. Data Science.

Sumário

Introdução	7
1 Apresentação dos dados	10
1.1 Processo de extração dos dados	10
1.2 Demografia da corrupção espanhola	13
1.3 Evolução temporal dos casos de corrupção espanhóis	19
2 Estrutura da rede de corrupção espanhola	22
2.1 Rede de corrupção espanhola	22
2.2 Vértices recorrentes e medidas de centralidade	27
2.3 Ataques direcionados por cargo e desmantelamento da rede	32
3 Dinâmica de crescimento da rede de corrupção espanhola	36
3.1 Visualizando a dinâmica da rede	36
3.2 Evolução temporal da distribuição de grau	37
3.3 Dinâmica das grandes componentes da rede	39
3.4 Estrutura de comunidades	44
4 Reincidência criminosa e um modelo para redes de corrupção	48
4.1 Dinâmica de reincidência	48
4.2 Comportamento do modelo computacional	51
4.3 Comparação com resultados empíricos	54
Conclusões e perspectivas	59
Apêndice A: Conceitos de ciência de redes	61
Apêndice B: Conceitos de estatística	66
Referências Bibliográficas	75

Introdução

O primeiro passo para resolver qualquer problema é defini-lo adequadamente. Entretanto, tratando-se de corrupção, existe uma dificuldade inerente ao tentar fornecer uma definição precisa. Embora as consequências de corrupção sejam visíveis e marcantes na sociedade, o conceito de corrupção é dinâmico e os agentes envolvidos procuram constantemente por novas maneiras de legitimar suas ações criminosas. Todavia, para fins práticos, tomaremos como corrupção atos de interesse pessoal que negligenciam a necessidade do Estado em ser imparcial [1] e práticas relacionadas ao uso indevido do poder incumbido [2].

Organizações criminosas e corrupção representam um problema transnacional [3] capaz de afetar profundamente o bem-estar coletivo da sociedade. Governos modernos têm por objetivo estabelecer e impor regras que garantam coesão social, liberdades individuais e bem-estar coletivo. Corrupção e crime organizado prejudicam esses objetivos criando estruturas sociopolíticas que capturam parte do Estado para sacrificar o bem-estar coletivo em prol do benefício de poucos.

Estimativas do Banco Mundial indicam que o custo anual da corrupção ultrapassa os 3 trilhões de dólares, sendo 1 trilhão de dólares vinculados ao pagamento de propinas [4]. Essas práticas dificultam o desenvolvimento da economia, desperdiçam recursos do Estado, limitam inovação [5], prejudicam o meio ambiente [6], impossibilitam distribuições justas de recursos [7] e aumentam a lacuna entre diferentes classes sociais [8]. Além disso, essas atividades geram desconfiança no Estado de Direito, promovem instabilidade política e fragilizam as instituições por meio de práticas como o nepotismo, impossibilitando que pessoas qualificadas ocupem cargos relevantes.

Outro aspecto imanente da corrupção é seu intrincado círculo vicioso. A existência de uma corrupção sistêmica afeta negativamente vários aspectos de um país e pode produzir cenários ainda mais propícios às atividades ilícitas. López-Iturriaga e Senz [9] comentam que a Espanha é um bom exemplo desse círculo vicioso. Pesquisas de opinião pública de 2017 sugerem que 94% dos cidadãos espanhóis enxergam corrupção como um problema generalizado [10]. Estimativas de 2018 indicam que os custos da corrupção espanhola ultrapassam 90 bilhões de euros anualmente, correspondendo a cerca de 8% do PIB do país [11]. A corrupção estava entre as maiores preocupações dos cidadãos espanhóis em 2019 [12] e, de acordo com o Índice de Percepção de Corrupção [13], o país possui uma

pontuação abaixo da média dos países da União Europeia.

A Espanha foi gravemente afetada pela crise de crédito global de 2008 [14]. O mercado imobiliário espanhol entrou em colapso, levando a uma recessão profunda que persistiu por vários anos. Sem medidas para promover o crescimento econômico, o país buscou um endurecimento da política fiscal, mas não conseguiu desvalorizar a taxa de câmbio. Os resultados da crise foram devastadores, incluindo uma forte desaceleração econômica, aumento severo no desemprego [15] e falência de muitas empresas [16]. A opinião pública sobre a situação foi agravada por diversos escândalos de corrupção que surgiram nesse período, vários deles associados aos dois maiores partidos políticos da Espanha: o Partido Popular (PP) e o Partido Socialista Operário Espanhol (PSOE). Esses partidos alternam o poder desde 1982 e compartilham o interesse em manter o cenário político e a estrutura econômica favorável a ambos. Similar ao que ocorre em muitos outros países, a Espanha apresenta uma quantidade significativa de escândalos de corrupção, tornando-se um importante exemplo para o estudo efetivo da corrupção e do crime organizado.

Embora a criminalidade seja onipresente na sociedade, organizações criminosas (como as envolvidas em corrupção) não operam de forma aleatória e tampouco uniforme no tempo e espaço. Contrariamente, o funcionamento dessas organizações depende de um sofisticado grau de coordenação entre os agentes criminosos, os quais formam estruturas de interação hierarquicamente geridas visando principalmente o lucro monetário. Essas associações entre os agentes são fundamentais para o sucesso das práticas criminosas de suas organizações, as quais dificilmente poderiam ser executadas por indivíduos isolados. Apesar da importância fundamental dessas inter-relações, a literatura científica sobre crime organizado e corrupção possui um forte viés econômico e sociológico [17]. Além disso, esses estudos se concentram em investigar os efeitos da ação de organizações criminosas por meio de índices econômicos e de percepção social em escala de países. Sabe-se muito pouco sobre os processos e padrões de operação de organizações criminosas implicadas em atividades ilícitas. Levando em conta seus aspectos não triviais, esse tema tem sido tratado com sucesso por abordagens científicas e objetivas advindas de ciência de redes e sistemas complexos [18–21].

Diferentemente de crimes menos sofisticados, as práticas de corrupção requerem uma intricada coordenação entre os indivíduos envolvidos para o sucesso da empreitada criminosa. Essa característica faz com que esses fenômenos sociais possuam um caráter sistêmico, marcado por respostas não lineares e comportamentos emergentes difíceis de prever e serem modelados a partir de abordagens tradicionais de estatística. Nesse sentido, a Física de Sistemas Complexos surge como um avanço para tratar esses fenômenos altamente adaptativos e não lineares.

Um sistema é considerado complexo não só porque possui uma estrutura intricada, mas também porque sua evolução temporal não é facilmente explicada como função do comportamento de seus constituintes isolados. A representação de sistema complexo que

trataremos será embasada em ciência de redes [22]. Nessa interpretação, dispomos de um conjunto de nós (ou vértices) representando objetos discretos. As ligações entre esses nós correspondem ao conjunto de relações entre esses elementos. Os nós e ligações podem indicar as mais variadas informações e as redes formadas são muito mais do que simples relações entre esses conjuntos; efeitos como não linearidade, emergência e adaptação são exemplos de possíveis características desses sistemas [23]. Redes sociais configuram um exemplo clássico na qual indivíduos são representados por nós e as ligações dependem de algum tipo de relação social.

No caso da corrupção, o constituinte fundamental é o indivíduo envolvido na prática criminosa. A associação entre esses agentes é essencial para as operações de organizações criminosas. Por consequência, o uso de redes complexas para representar essas organizações ligadas à corrupção é bastante natural e consistente: os vértices representam criminosos ou suspeitos e as conexões indicam as parcerias entre eles. Dessa forma, os envolvidos em um caso de corrupção estão conectados entre si e aqueles que participam em mais de um escândalo atuam como elos entre os diferentes casos de corrupção.

Combinado com dados de qualidade, esse tipo de representação permite investigar padrões e impactar profundamente o conhecimento sobre o funcionamento do crime de corrupção. Tal qual vestígios de uma cena criminal, os traços e padrões desvendados por essas análises baseadas em redes podem ser usados para prever, fragilizar ou até mesmo interromper o funcionamento dessas organizações criminosas. Embora seja inovadora e recente, esse tipo de abordagem apresenta precedentes de sucesso. Estudos anteriores baseados em uma rede de corrupção brasileira mostraram que é possível prever novos envolvimentos com até 25% de acerto [18]. Outros pesquisadores também investigaram diferentes aspectos ligados ao desmantelamento e controle de redes criminosas brasileiras [21]. Esses estudos demonstram a grande capacidade e adequação de métodos baseados em sistemas complexos e ciência de redes para compreender o funcionamento de organizações criminosas envolvidas em corrupção.

Nesta dissertação, seguimos uma abordagem similar para investigar uma rede composta por indivíduos envolvidos em casos de corrupção da Espanha. Este trabalho está organizado da seguinte maneira. No Capítulo 1 apresentamos o processo de extração dos dados sobre os casos de corrupção e algumas análises demográficas. O Capítulo 2 apresenta uma análise empírica de diversos aspectos estruturais da rede de corrupção espanhola. Em seguida, no Capítulo 3, passamos a considerar o caráter temporal dessa rede de corrupção. Motivados pelos resultados empíricos, no Capítulo 4, propomos um modelo computacional para a formação e evolução de redes de corrupção e comparamos os resultados com o caso empírico. Encerramos essa dissertação com um resumo das principais contribuições e com nossas perspectivas para pesquisas futuras. Os Apêndices A e B apresentam alguns métodos de ciência de redes e conceitos de estatística utilizados em nossas investigações.

Capítulo 1

Apresentação dos dados

Embora a quantidade de informações disponibilizadas na Internet seja exorbitante e cresça em um ritmo sem precedente, ainda existe uma dificuldade em se obter dados precisos relacionados ao funcionamento de organizações criminosas envolvidas em corrupção. Parte dessa dificuldade decorre do esforço dos agentes envolvidos em manter as atividades ilegais ocultas e ao fato de que essas informações podem conter vieses de divulgação dos meios de comunicação. Essas particularidades resultam em uma escassez de estudos voltados à compreensão empírica dos processos de corrupção na escala do indivíduo implicado nesse tipo de crime.

Apesar de possíveis ressalvas, jornais e revistas de grande circulação oferecem a maior parte das informações públicas sobre corrupção. Sendo assim, esses meios de comunicação podem ser utilizados como base de dados para estudos sobre corrupção. Neste capítulo, apresentamos uma base de dados sobre casos de corrupção da Espanha obtida da página www.casos-aislados.com. Esses dados compõem a principal fonte de informação usada em nossas investigações. Por isso, este capítulo também apresenta uma análise demográfica desses dados.

1.1 Processo de extração dos dados

A página www.casos-aislados.com apresenta uma listagem dos casos de corrupção da Espanha. Essa página não está vinculada a partidos políticos ou organizações privadas e tem por objetivo fornecer informações atualizadas sobre os escândalos de corrupção espanhóis. Atualmente, a página contém informações sobre mais de 500 casos de corrupção. Para extrair essas informações de forma automatizada, utilizamos a biblioteca *Scrapy* [24] da linguagem de programação *Python* [25].

De modo mais específico, escrevemos um primeiro algoritmo usando a *Scrapy* (também denominado por *crawler* ou *spider*) para coletar as *urls* das subpáginas do site *casos-aislados* contendo os detalhes dos casos de corrupção. A Figura 1.1 mostra esse código.

```

1 import scrapy
2 from casos_aislados.items import CasosAisladosItem
3
4 class GetLinksSpider(scrapy.Spider):
5     name = "links"
6
7     allowed_domains = ['https://www.casos-aislados.com']
8     start_urls = ['https://www.casos-aislados.com/en/tramas.php']
9
10    def parse(self, response):
11        for sel in response.xpath("/html/body/div[2]/div/div[4]/div")[2:]:
12
13            item = CasosAisladosItem()
14            item['cases_link'] = "https://www.casos-aislados.com/en/" + str(sel.xpath("div[1]/a/@href").extract()[0])
15
16            yield item

```

Figura 1.1: Algoritmo usado para obter as *urls* de todas as subpáginas do site *casos-aislados* contendo as informações sobre os escândalos de corrupção da Espanha.

Essas *urls* foram armazenadas em um arquivo de texto para serem usadas posteriormente. Em seguida, escrevemos um segundo algoritmo para coletar as informações relevantes sobre os escândalos detalhadas em cada subpágina do site *casos-aislados*. A Figura 1.2 mostra uma parte desse código para fins ilustrativos.

```

1 import scrapy
2 from casos_aislados.items import CasosAisladosItem
3
4 urls = list()
5 with open("links_casos.csv", "r") as f:
6     for line in f.readlines():
7         urls.append(line.replace('\n', ''))
8
9 class GetDataSpider(scrapy.Spider):
10    name = "cases"
11    allowed_domains = ["https://www.casos-aislados.com"]
12    start_urls = urls
13
14    def parse(self, response):
15        item = CasosAisladosItem()
16        regiao = response.xpath("/html/body/div[2]/div[1]/div[4]/div[3]/div[1]/div[2]/ul/li[2]/span/text()").extract()
17        partido = response.xpath("/html/body/div[2]/div[1]/div[4]/div[3]/div[1]/div[2]/ul/li[1]/span/text()").extract()
18        nome_caso = response.xpath("/html/body/div[2]/div[1]/div[4]/div[1]/text()").extract()
19        ano_evento = response.xpath("/html/body/div[2]/div[1]/div[4]/div[3]/div[1]/div[2]/ul/li[5]/span/text()").extract()
20
21        if len(regiao) > 0:
22            item['regiao'] = regiao[0]
23
24        if len(partido) > 0:
25            item['partido'] = partido[0]
26
27        if len(nome_caso) > 0:
28            item['nome_caso'] = nome_caso[0].strip(' \t\n').replace("Sobre el Caso ", "")
29
30        if len(ano_evento) > 0:
31            item['ano_evento'] = ano_evento[0]
32
33        i, j = 2, 3
34        dados_envolvimento = dict()
35        while True:
36
37            item['dados_envolvimento'] = dados_envolvimento
38            yield item

```

Figura 1.2: Algoritmo usado para obter as informações relevantes de cada subpágina do site *casos-aislados* relacionada aos escândalos de corrupção da Espanha.

As informações obtidas contêm o nome do escândalo de corrupção, o ano no qual os eventos iniciaram e as regiões geográficas dos acontecimentos. Além disso, também extraímos os nomes dos envolvidos, seus respectivos cargos e os partidos ou organizações associadas a essas pessoas. Realizamos um pré-processamento detalhado para remover informações irrelevantes e dados inválidos. Além disso, fizemos uma verificação das informações dos principais escândalos de corrupção, comparando-as com as informações de

outros meios de comunicação (principalmente do jornal espanhol *El Mundo*). Inspecionamos ainda os nomes dos indivíduos para verificar a existência de nomes muito semelhantes ou idênticos. Ademais, acentos e outros caracteres especiais foram removidos com a finalidade de deixar os dados no formato de texto plano.

Em algumas poucas ocorrências, o ano de início dos eventos de certos escândalos de corrupção não estavam informados no site *casos-aislados*. Nesses casos, realizamos uma busca cuidadosa em páginas da Internet e completamos as informações ausentes. A Tabela 1.1 resume a estrutura das informações obtidas após o pré-processamento dos dados.

	Ano	Região	Dados dos envolvidos	Nome do escândalo	Partidos/organizações
0	1989	Comunidad de Madrid	[envolvidos, cargos]	Gürtel/Primera Época	PP
1	1990	País Vasco	[envolvidos, cargos]	Hacienda de Irún	PNV
2	1990	Cataluña	[envolvidos, cargos]	Treball	CiU
3	1991	País Vasco	[envolvidos, cargos]	Pagalday	PNV
4	1992	Galicia	[envolvidos, cargos]	PP de Valga	PP
5	1993	Comunidad de Madrid	[envolvidos, cargos]	Banesto	Bancos
6	1993	Comunidad Valenciana	[envolvidos, cargos]	Carmelitas	PP
7	1994	Navarra	[envolvidos, cargos]	Urralburu	PSOE
8	1995	La Rioja	[envolvidos, cargos]	Casalarreina	PP
9	1995	Cataluña	[envolvidos, cargos]	Hacienda	PSOE
:	:	:	:	:	:
430	2017	Comunidad de Madrid	[envolvidos, cargos]	Rosario	Empresas
431	2017	Región de Murcia	[envolvidos, cargos]	Soler	Cs
432	2017	Comunidad Valenciana	[envolvidos, cargos]	À Punt	Empresas Públicas
433	2017	Comunidad Valenciana	[envolvidos, cargos]	Cártel del fuego	PP
434	2017	Galicia	[envolvidos, cargos]	Cebra	PP
435	2018	Comunidad Valenciana	[envolvidos, cargos]	Brugger	PP
436	2018	Comunidad Valenciana	[envolvidos, cargos]	Maratón Santa Pola	PSOE

Tabela 1.1: Resumo das informações coletadas do site *casos-aislados*.

As informações que coletamos abrangem 437 escândalos de corrupção que ocorreram entre 1989 e 2018 (30 anos), totalizando 2753 pessoas envolvidas. Esses dados possuem números consideravelmente maiores quando comparados a um estudo anterior sobre corrupção brasileira reportado em 2018 na referência [18]. Nesse outro estudo, foram investigados 65 casos de corrupção ocorridos no Brasil entre 1987 e 2014, com 404 pessoas envolvidas.

Apesar de nosso esforço para tornar o presente conjunto de dados confiável, entendemos que a simples menção em um escândalo de corrupção não garante que uma pessoa seja considerada culpada e condenada judicialmente. Os procedimentos judiciais relacionados a grandes escândalos de corrupção podem levar anos ou até mesmo décadas, e alguns nunca chegam a um veredito final. Sendo assim, por questões legais, os nomes de todos os envolvidos foram anonimizados e não são apresentados em nosso trabalho. Além disso, é provável que algumas pessoas envolvidas em escândalos de corrupção não tenham sido descobertas. Consequentemente, não é possível afirmar que todos os indivíduos envolvidos em um dado escândalo de corrupção foram identificados durante as investigações e, nesse sentido, nossos dados podem estar incompletos. Infelizmente, essas limitações são inerentes a esses tipos de informações. Entretanto, o foco principal desse estudo está relacionado a padrões gerais dos processos de corrupção e, portanto, essas possíveis imperfeições dos dados devem impactar minimamente nossos achados empíricos. De fato, nosso trabalho irá mostrar que certos padrões relacionados ao crime de corrupção parecem não depender de especificidades menores dos dados.

1.2 Demografia da corrupção espanhola

A Espanha é organizada em dezessete comunidades autônomas e duas cidades autônomas (Ceuta e Melilla). A Figura 1.3 mostra o mapa da Espanha com relação ao número de envolvidos em escândalos de corrupção em suas regiões. Os três locais que possuem as maiores quantidades de implicados são Andaluzia (616 pessoas), Comunidade de Madri (512 pessoas) e Comunidade Valenciana (428 pessoas). Vale observar que embora algumas regiões concentrem a maior parte dos envolvidos em casos de corrupção, nenhuma localidade é desprovida desse tipo de crime.

Nessa mesma figura, o gráfico à esquerda representa a distribuição acumulada (complementar) do número de envolvidos por região da Espanha. A escala logarítmica no eixo y usada nesse gráfico sugere um comportamento exponencial para essa distribuição. Esse comportamento exponencial indica a existência de uma escala típica para o número de pessoas envolvidas. De fato, ao ajustar um modelo exponencial a esses dados, encontramos um bom acordo e um número característico de pessoas aproximadamente igual a 146.

Outra maneira de visualizar a quantidade de corrupção em cada região é por meio do gráfico de barras da Figura 1.4. Essa representação ressalta melhor a participação de todas as regiões da Espanha. Além disso, essa figura apresenta o aspecto temporal do surgimento dos escândalos por meio do código de cores usado nas barras coloridas. Podemos verificar que a ocorrência de casos de corrupção não está uniformemente distribuída no tempo na maioria das regiões espanholas. Esse é o caso, por exemplo, da região de Castela-Mancha que apresenta apenas escândalos mais recentes.

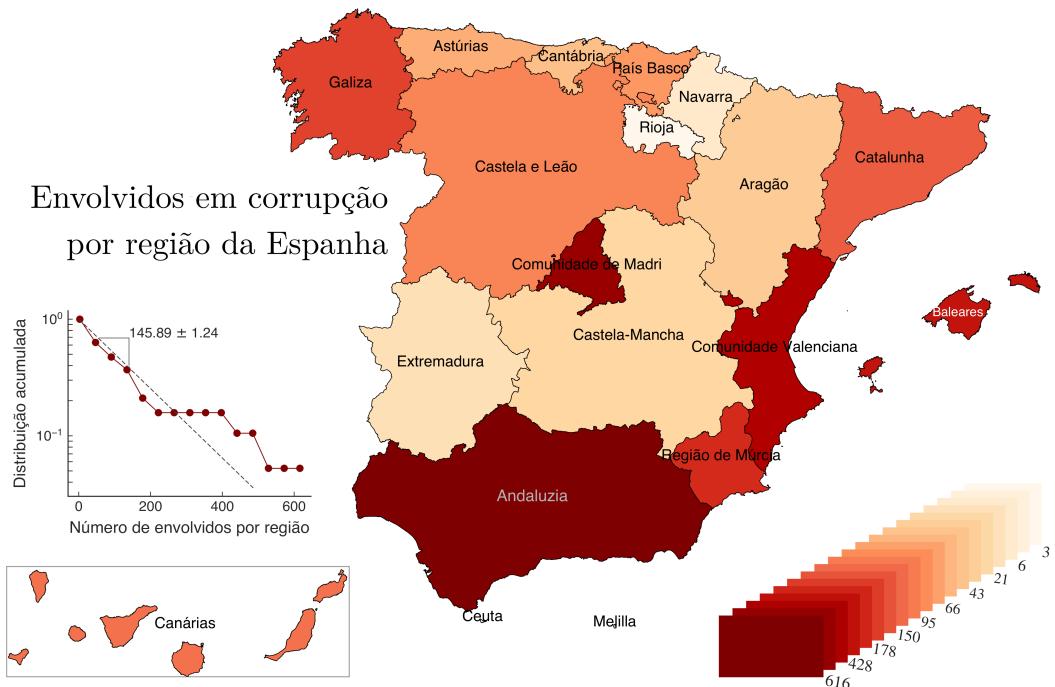


Figura 1.3: Mapa da Espanha exibindo o número total de envolvidos em corrupção por região. A inserção à esquerda mostra a distribuição acumulada complementar em escala mono-logarítmica do número de envolvidos em cada região. O comportamento linear nessa escala sugere uma distribuição exponencial para o número de envolvidos por região. A linha tracejada mostra um ajuste do modelo exponencial. O parâmetro associado ao decaimento da curva aponta que o número típico de envolvidos por região é aproximadamente igual a 146.

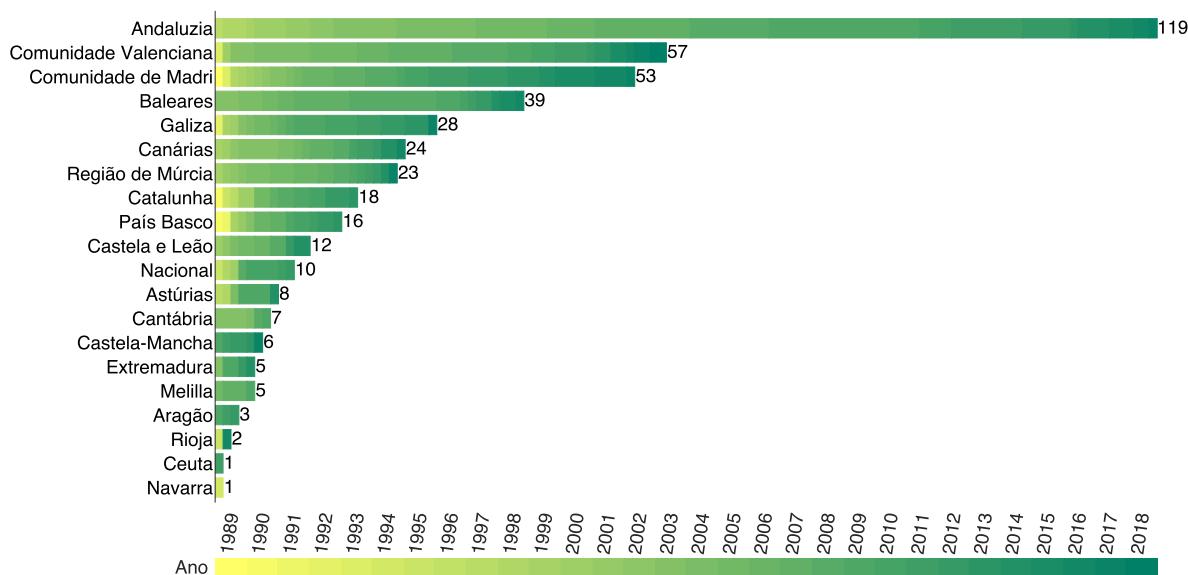


Figura 1.4: Número de escândalos políticos por região da Espanha. Casos que abrangem o país todo estão contabilizados como “Nacional” e o código de cores se refere à evolução temporal do número de casos por ano.

Podemos associar cada escândalo de corrupção aos partidos políticos espanhóis, empresas e organizações em geral. Para isso, consideramos que um caso está vinculado a determinada organização quando um ou mais de seus componentes possuem relações com essas organizações. A Figura 1.5 mostra o número de casos por organização e ilustra o aspecto temporal da ocorrência de casos por meio do código de cores das barras. Podemos notar que os dois maiores partidos espanhóis lideram em número de escândalos. Em 2019, essas duas organizações políticas ocupavam aproximadamente 60% dos assentos parlamentares espanhóis [26] e, do total de 2753 indivíduos potencialmente envolvidos em ações corruptas, o Partido Popular (PP) e o Partido Socialista Operário Espanhol (PSOE) representam, respectivamente, 56% e 21% desse total de envolvidos. Além disso, juntos, esses dois partidos estão associados a 72,5% de todos os 437 casos.

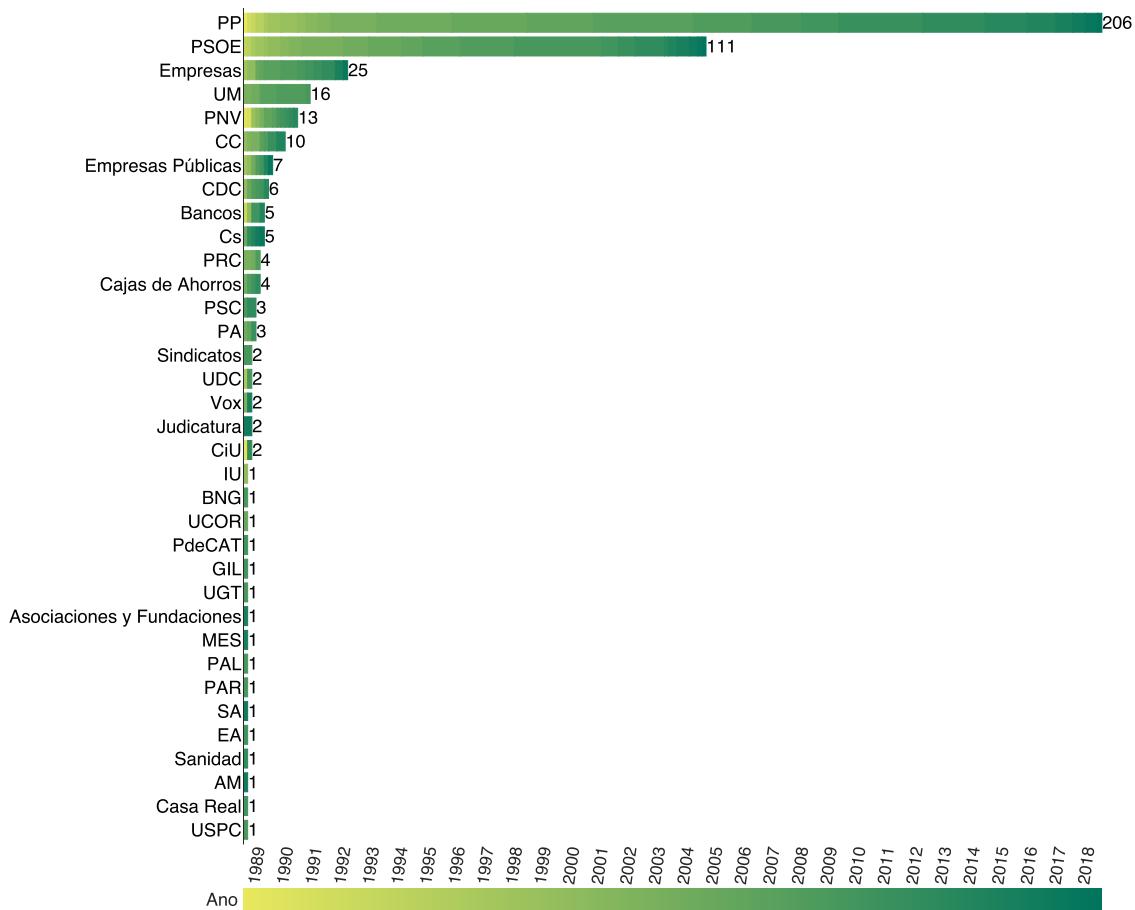


Figura 1.5: Número de escândalos políticos separados por partidos e organizações. Notamos que os dois maiores partidos espanhóis estão presentes na maioria dos casos de corrupção. Vale observar ainda que o Partido Popular participou em quase o dobro de casos do que o Partido Socialista Operário Espanhol.

De fato, é possível afirmar que algumas das maiores tramas de corrupção política da Espanha estão associadas a seus dois maiores partidos políticos. Essas tramas se desdobram em diversos casos de corrupção nas diversas regiões espanholas e apresentam suas

especificidades. O caso Gürtel, por exemplo, é um dos maiores escândalos de corrupção que o país possui, cobrindo vários tipos de acusações e centenas de implicados. Esse caso refere-se a um gigantesco esquema de suborno de contratos envolvendo empresários espanhóis e líderes locais do Partido Popular. Esse esquema de corrupção esteve concentrado inicialmente em Madrid e Valência, mas se espalhou por toda Espanha.

Associado ao Partido Socialista Operário Espanhol, destacamos outro mega-escândalo de corrupção espanhol conhecido como o caso ERE. Persistindo por cerca de uma década e se desdobrando em diversos casos na região de Andaluzia, essa trama envolveu a malversação de centenas de milhões de euros de ajuda pública provenientes de um fundo para desempregados e empresas em crise. O acrônimo ERE advém de *Expediente de Regularización de Empleo*, ou Arquivos de Regulamento do Emprego, um procedimento legal para financiar demissões coletivas de trabalhadores por empresas em dificuldade. No entanto, investigações mostraram que esse fundo foi extensivamente mal alocado e canalizado para pessoas que não eram abrangidas pelo programa.

Outra trama de corrupção que teve grande repercussão na Espanha foi o caso Púnica. Esse caso representa um esquema de propinas em contratos governamentais envolvendo líderes políticos locais, empresários espanhóis e, principalmente, membros do Partido Popular da Espanha. Afetando principalmente as comunidades autônomas de Madri, Região de Múrcia, Castela e Leão e Valência, a investigação desse escândalo envolveu diversos indivíduos, localidades espanholas e se desmembrou em vários casos.

Esses três exemplos estão entre os mais notórios dos muitos escândalos que apresentam diversos desdobramentos. A Figura 1.6 mostra os dez maiores casos de corrupção em relação ao total de envolvidos com seus respectivos números de casos desdobrados. Podemos notar que mais de 80% dessas tramas que produziram desdobramentos estão associadas aos dois maiores partidos políticos espanhóis, representados pelas cores azul e vermelha. Além disso, uma vez que esses dois partidos estão associados a mais de 72% de todos os casos dos nossos dados, podemos explorar apenas suas contribuições para a quantidade de escândalos por comunidades autônomas. A Figura 1.7 mostra o número de casos por localidade divididos entre esses dois partidos. Observamos uma diferença clara na forma pela qual esses escândalos se distribuem na Espanha. O PSOE possui uma concentração de casos na região de Andaluzia (parcialmente devido aos desdobramentos do ERE), enquanto o PP tem seus casos distribuídos de forma menos concentrada (destacando-se em Valenciana e Madri).

Analisamos também o número total de envolvidos por região associados aos dois partidos. Conforme mostra a Figura 1.8, notamos que o Partido Popular tem um número de envolvidos quase três vezes maior do que o PSOE (1568 versus 599). Essa figura também mostra que o PSOE tem seus envolvidos mais concentrados em Andaluzia, enquanto os implicados do PP estão mais distribuídos pelas regiões da Espanha. De forma geral, nossos dados indicam que os partidos PP e PSOE têm papel central na corrupção espanhola.

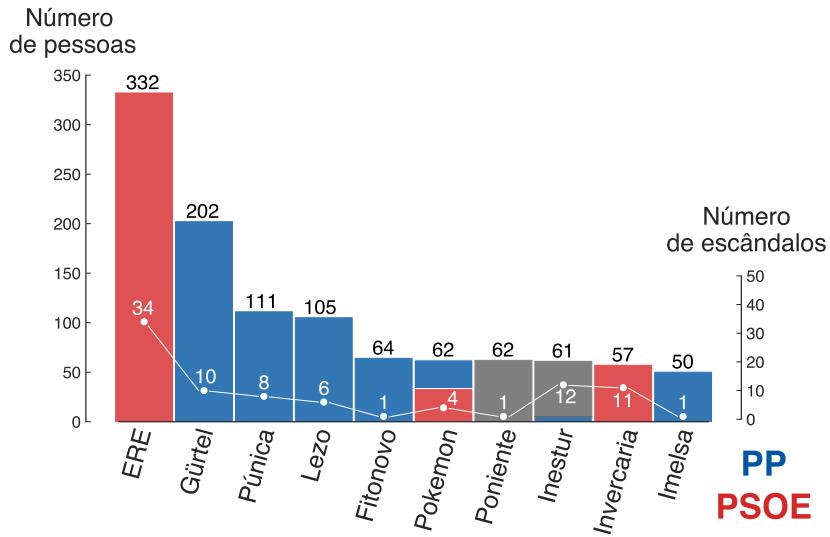


Figura 1.6: Dez maiores casos de corrupção em número de implicados e seus desdobramentos. O caso Poniente está completamente associado ao Partido de Almería (PAL) enquanto Inestur integra *Unió Mallorquina* (UM, 11 dos 12) e Partido Popular (PP, 1 dos 12).

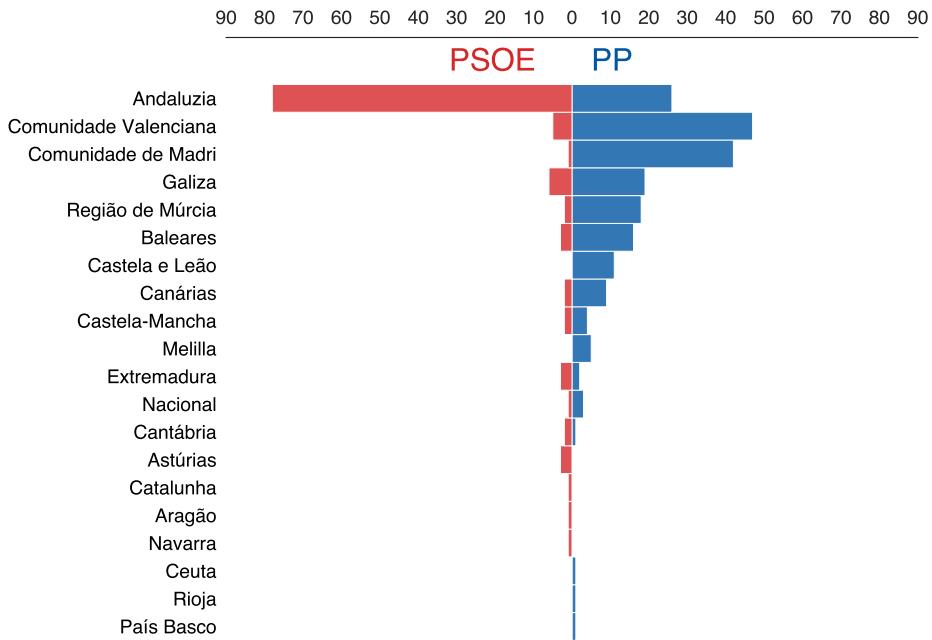


Figura 1.7: Total de escândalos associados aos partidos PP e PSOE para as diferentes regiões espanholas. O Partido Popular esteve envolvido em um total de 206 escândalos, enquanto o PSOE esteve associado a 111 casos. Notamos que os casos que envolveram o PP estão mais distribuídos pela Espanha.

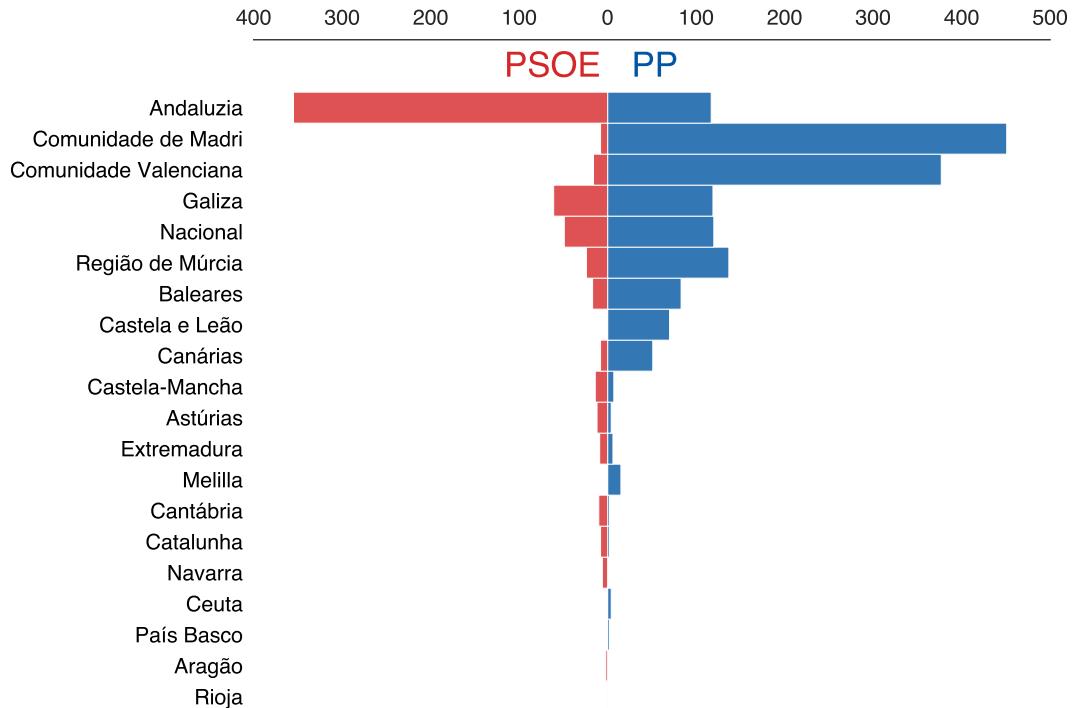


Figura 1.8: Total de envolvidos por região da Espanha associados aos partidos PSOE e PP. Existem 1568 pessoas envolvidas em casos de corrupção associadas ao partido PP. Esse número é 599 para o partido PSOE.

Outro aspecto interessante para analisarmos se refere ao número de homens e mulheres envolvidos em corrupção. Em nosso estudo, os gêneros foram obtidos utilizando um algoritmo em *Python* chamado *genderator* [27]. Esse procedimento se baseia no nome e sobrenome do indivíduo e segue como parâmetro o *Instituto Nacional de Estadística* [28], além de retornar a respectiva confiança associada ao resultado. Somado a isso, utilizamos informações adicionais de pesquisas *online* para obter os gêneros de todos os envolvidos. A Figura 1.9 mostra o total de implicados separados por sexo para as diferentes regiões da Espanha. Podemos observar que o número de homens envolvidos supera o de mulheres em todas as localidades. Esse padrão faz com que o número total de homens envolvidos em corrupção (2173) seja muito superior ao de mulheres (556).

A menor participação feminina em casos de corrupção da Espanha está de acordo com outros trabalhos de criminologia [29–31]. Esses estudos mostram que mulheres cometem menos crimes do que homens e apontam que parte desse comportamento pode ser relacionado a aspectos evolucionários do gênero masculino (tendência de correr mais riscos e comportamentos violentos), além de questões sociais ligadas à desigualdade de gênero.

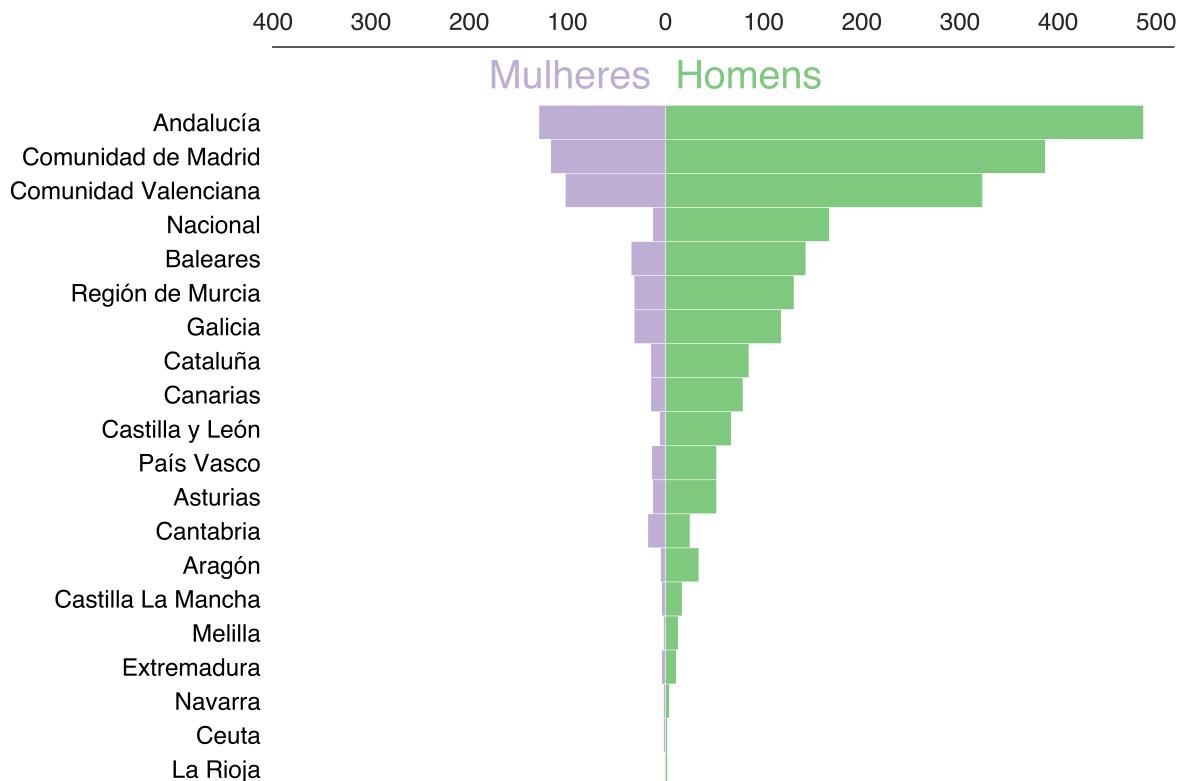


Figura 1.9: Número de pessoas envolvidas em casos de corrupção agrupadas por gênero e região.

1.3 Evolução temporal dos casos de corrupção espanhóis

Para investigar aspectos sobre a evolução temporal da corrupção na Espanha, estimamos a quantidade de implicados e casos para todos os anos entre 1989 e 2018. Os resultados das Figuras 1.10A e 1.10B indicam que não há uma tendência clara de aumento no número de envolvidos nem no número de casos. No entanto, é perceptível que essas quantidades apresentam uma alta na atividade entre 2003 e 2013 e um máximo ao redor de 2009. Esse comportamento se contrasta com o estudo sobre corrupção no Brasil [18], no qual o número de pessoas envolvidas apresenta uma ligeira tendência de crescimento no tempo.

Analisamos também a distribuição acumulada (complementar) do tamanho dos escândalos espanhóis (no quesito número de envolvidos). Conforme mostra a Figura 1.10C, essa distribuição pode ser ajustada por uma distribuição exponencial, com o número típico de pessoas aproximadamente igual a 7. Assim, embora existam escândalos envolvendo um grande número de pessoas, o tamanho típico de escândalos de corrupção da Espanha é re-

lativamente pequeno e similar ao encontrado para a corrupção brasileira [18]. Entretanto, a curva exponencial subestima o número de envolvidos nos grandes casos de corrupção espanhóis. Ainda assim, o modelo exponencial descreve razoavelmente bem uma grande parcela dos dados. Vale observar que apenas 20% de todos os casos de corrupção espanhóis possuem mais de 10 indivíduos, um comportamento parecido ao observado para o Brasil (no qual 17% dos casos envolvem mais de 10 pessoas).

Esses resultados reforçam a hipótese levantada na referência [18] de que as pessoas tendem a agir em pequenos grupos quando envolvidas em processos criminosos, provavelmente pela maior facilidade em manter a atividade ilícita oculta.

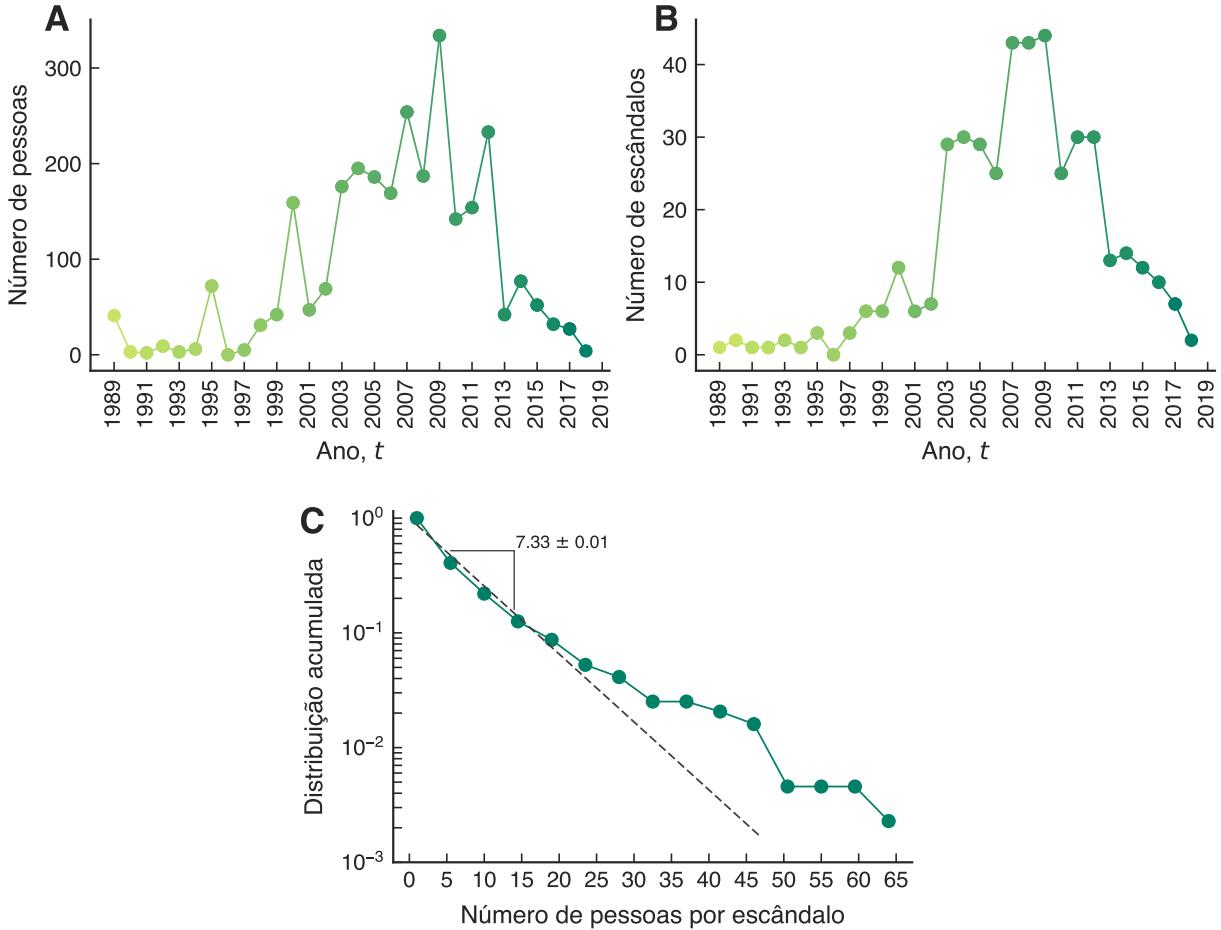


Figura 1.10: Evolução temporal do (A) número de envolvidos e do (B) número de casos ao longo dos 30 anos cobertos por nossa base de dados. (C) Distribuição acumulada complementar do número de implicados nos casos de corrupção. A linha tracejada representa uma distribuição exponencial ajustada aos dados via método de máxima verossimilhança [32]. Esse ajuste indica que o número típico de envolvidos em um caso de corrupção é igual a 7.33 ± 0.01 .

Uma vez que os dois maiores partidos espanhóis compõem a maioria dos casos de corrupção, é interessante analisar a evolução temporal do número acumulado de casos e implicados considerando apenas as informações referentes a esses dois partidos. Essa análise é mostrada nas Figuras 1.11A e 1.11C, enquanto as Figuras 1.11B e 1.11D exibem

o número anual dessas quantidades e os valores médios ao longo dos intervalos de tempo em que cada partido esteve no poder (ou seja, o Primeiro Ministro era do partido).

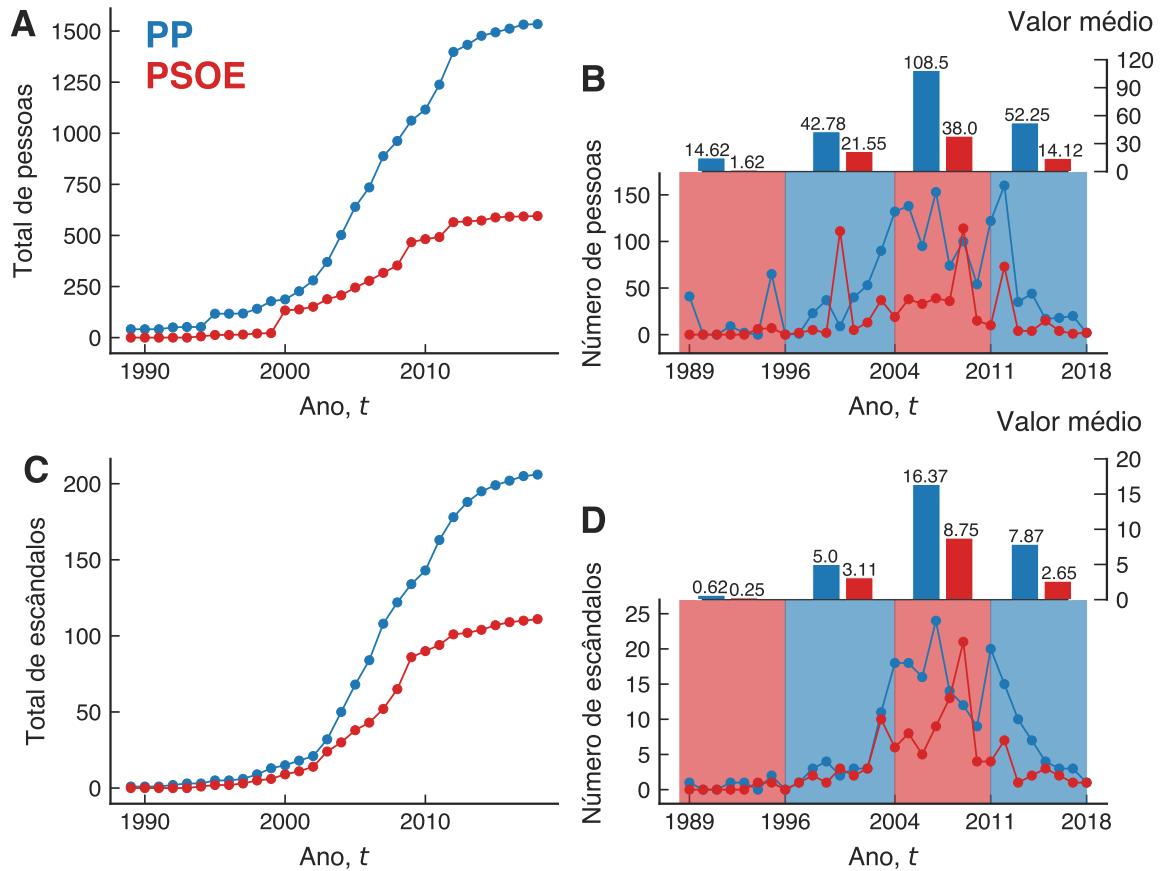


Figura 1.11: Contribuição dos partidos PP e PSOE nas séries temporais de escândalos de corrupção e seus envolvidos. (A) Quantidade acumulada de pessoas envolvidas em função do tempo. (B) Número anual de implicados e valor médio durante os intervalos referentes às mudanças de poder político na Espanha. (C) Quantidade acumulada de casos de corrupção ao longo do tempo. (D) Número anual de casos e valor médio dessa quantidade durante os intervalos referentes às mudanças de poder político na Espanha.

Diferentemente dos resultados reportados para o Brasil [18], as variações no número de casos de corrupção e de envolvidos não parece estar diretamente relacionada à alternância de poder entre os dois partidos. Conforme especulado no estudo brasileiro, mudanças repentinas (nos números de casos e pessoas) coincidentes com alternância de poder poderiam indicar o direcionamento de investigações para os partidos de oposição ao governo. Porém, este não parece ser o caso da Espanha.

Capítulo 2

Estrutura da rede de corrupção espanhola

Neste capítulo, apresentamos uma análise descritiva da rede de corrupção política da Espanha. Iniciamos pelo processo de construção dessa rede ao levar em conta todos os casos de corrupção em nossa base de dados. Em seguida, descrevemos diversos aspectos estáticos dessa rede relacionados a padrões demográficos e medidas de centralidade. Por fim, apresentamos uma análise sobre o desmantelamento dessa rede.

2.1 Rede de corrupção espanhola

Uma rede complexa (ou grafo) é um objeto composto por um conjunto de vértices ou nós e um conjunto de arestas ou ligações¹. No caso da rede de corrupção política da Espanha, vamos considerar as pessoas envolvidas como vértices da rede. As ligações entre esses vértices indicam duas pessoas que participaram em um mesmo escândalo de corrupção.

Nessa análise inicial, não levamos em conta o aspecto temporal dos casos de corrupção e consideramos todos os 437 escândalos ocorridos entre 1989 e 2018 cobertos por nossa base de dados. A Figura 2.1 mostra uma visualização dessa rede composta por 2753 vértices e 27545 ligações. Nessa figura, as cores dos vértices indicam as diferentes organizações associadas aos indivíduos da rede. Para deixar a visualização mais clara, escolhemos as cores azul, vermelho e laranja para denotar (respectivamente) os partidos PP, PSOE e UM. Os vértices em preto indicam pessoas relacionadas à empresas privadas e os em cinza indicam pessoas associadas às demais organizações presentes em nossa base de dados (PNV, CC, Vox, empresas públicas, entre outras). Uma vez que pessoas ligadas aos partidos PP, PSOE, UM e empresas privadas correspondem a mais de 90% de todos os

¹Apresentamos uma visão geral dos principais aspectos de redes usados neste trabalho no Apêndice A.

envolvidos nos casos de corrupção, essa figura exibe o arranjo das principais organizações da rede de corrupção espanhola.

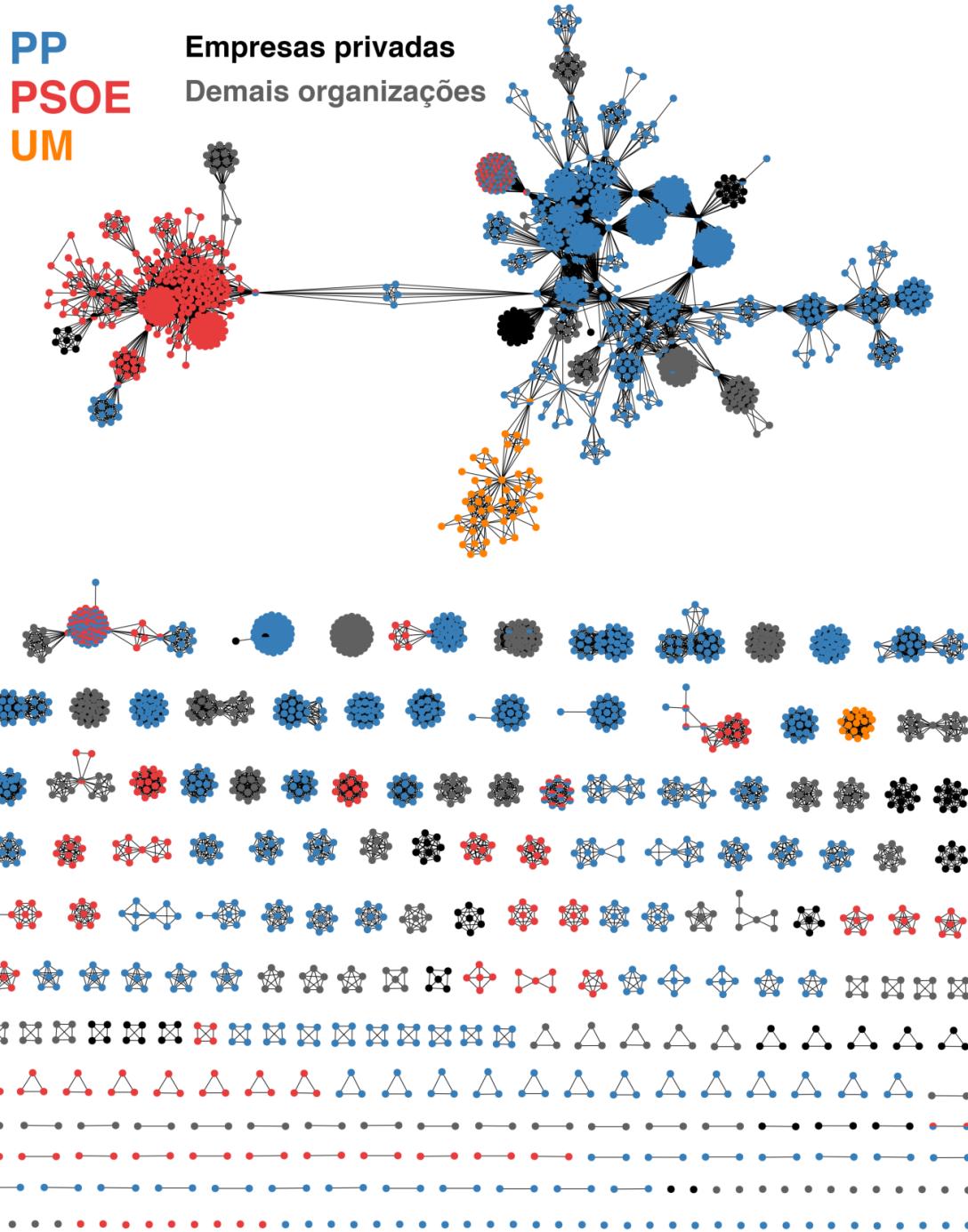


Figura 2.1: Visualização da rede de corrupção espanhola ao considerar todos os casos de corrupção ocorridos entre 1989 e 2018 em nossa base de dados. Cada vértice representa uma pessoa e as ligações indicam dois indivíduos que participaram, ao menos uma vez, em um mesmo caso de corrupção. Conforme indicado na legenda, as cores de cada vértice se referem às organizações relacionadas aos envolvidos nos casos de corrupção. Os poucos indivíduos que estiveram associados a mais de uma organização possuem vértices divididos igualmente por cores de cada organização.

Além de 58 envolvidos isolados, a rede de corrupção espanhola apresenta 197 componentes conectadas (organizadas em ordem de tamanho na Figura 2.1). A maior componente possui 1089 vértices e 14502 ligações, representando 40% de todos vértices e 53% de todas as ligações da rede. Essas porcentagens são menores quando comparadas à rede de corrupção brasileira investigada na referência [18], na qual a maior componente compreende 77% dos vértices e 93% das ligações de toda rede. Parte dessa diferença pode ser explicada pelo fato da rede brasileira possuir apenas 14 componentes e um número muito menor de envolvidos. Além disso, a rede de corrupção espanhola apresenta muitos casos pequenos e isolados, o que contribui para uma redução do número de conexões da rede.

É relativamente comum encontrar redes reais (principalmente de origem social) que não são inteiramente conectadas. Nesses casos, não existe um caminho que conecta dois vértices quaisquer. Esse fenômeno também ocorre na rede de corrupção espanhola mostrada na Figura 2.1. Em situações como essa é comum analisar a estrutura e dinâmica da maior componente da rede. No nosso caso, a maior componente contém os escândalos de corrupção mais importantes da Espanha e, portanto, a usaremos para extrair algumas informações estruturais.

O aspecto visual da maior componente da rede de corrupção espanhola evidencia uma polarização entre os envolvidos associados aos partidos PP e PSOE. Existe uma tendência muito maior de envolvimento em casos de corrupção entre indivíduos associados ao mesmo partido, com pouquíssimas conexões entre envolvidos associados a partidos diferentes. Entretanto, uma dessas conexões é responsável por unir as duas partes polarizadas da rede. Mais adiante, na Seção 3.3 do Capítulo 3, apresentamos uma descrição mais detalhada do processo de coalescência de componentes da rede, destacando os principais escândalos responsáveis pela estrutura mais conectada da rede de corrupção espanhola.

Quanto aos aspectos estruturais, obtemos um coeficiente de agrupamento² igual a 0.908 para a rede completa e 0.939 para a maior componente. Esses valores são bem próximos aos reportados para a rede de corrupção brasileira (0.925 e 0.929, respectivamente). Outra propriedade interessante de uma rede é sua homofilia ou assortatividade, a qual pode ser medida pelo coeficiente de assortatividade³. Esse coeficiente é igual a 0.74 para a rede toda e 0.59 para a componente gigante. O caso da rede de corrupção brasileira resulta em valores ligeiramente menores: 0.60 para rede completa e 0.53 para a maior componente [18]. Nos dois casos, esses valores relativamente altos de assortatividade indicam uma tendência maior de conexão entre vértices com grau similar. Em parte, esse comportamento pode ser explicado pela maneira como essas redes são construídas, ou seja, ao fato dos envolvidos serem incluídos na rede na forma de módulos completamente conectados (os escândalos) nos quais todos os vértices não recorrentes têm o mesmo grau.

Analisamos também a propriedade de mundo pequeno da rede de corrupção espanhola

²O Apêndice A.2.1 apresenta detalhes sobre o coeficiente de agrupamento.

³O Apêndice A.2.2 apresenta detalhes sobre o coeficiente de assortatividade.

calculando a média do caminho mais curto para a maior componente. Encontramos que a distância média entre os vértices da rede é de 5.11 passos, valor que é aproximadamente igual ao dobro do obtido para uma rede aleatória⁴ do mesmo tamanho e número de ligações (2.511 ± 0.003 passos, valor médio em 1000 realizações). Comparado ao estudo sobre redes de corrupção brasileira, esse valor médio para o caminho mais curto é razoavelmente maior (2.99 no caso brasileiro). Esse resultado sugere que, apesar dos vértices estarem relativamente próximos uns dos outros, os agentes corruptos espanhóis aumentaram sua distância com o crescimento da rede, comportamento possivelmente relacionado à polarização entre os partidos. Esse distanciamento maior entre os agentes da rede espanhola pode ser útil às organizações criminosas pois aumenta a separação entre as partes polarizadas da rede e contribui para que os escândalos fiquem mais isolados. Esse padrão também pode ajudar a explicar as diferenças entre os casos de corrupção brasileiros e espanhóis reportados nas Figuras 1.11B e 1.11D. Ou seja, a maior separação observada na rede espanhola pode dificultar possíveis investigações desencadeadas pela alternância do poder político entre os partidos. No geral, essas informações reforçam a ideia de que a evolução de redes criminosas visa a maximizar a ocultação de seus envolvidos, tal qual descrito pela teoria das sociedades secretas [33]. Para fins de comparação, a Tabela 2.1 apresenta os valores das medidas citadas anteriormente para as redes de ambos os países.

		Rede brasileira	Rede espanhola
Agrupamento	Rede Completa	0.925	0.908
	Maior Componente	0.929	0.939
Assortatividade	Rede Completa	0.60	0.74
	Maior Componente	0.53	0.59
Caminho médio	Maior Componente	2.99 (2.146 ± 0.002)	5.11 (2.551 ± 0.003)

Tabela 2.1: Valores de medidas estruturais das redes de corrupção brasileira e espanhola.

Na tentativa de deixar mais clara a separação entre os partidos políticos da Espanha, construímos uma rede formada pelos casos de corrupção. Nessa representação, cada vértice é um caso de corrupção e dois escândalos estão conectados se compartilham ao menos um envolvido. A Figura 2.2 mostra essa rede composta de 437 vértices e 750 ligações, na qual as cores dos vértices correspondem aos partidos e organizações associadas aos respectivos escândalos. Escândalos compostos por mais de uma organização são representados por vértices divididos igualmente nas cores de cada organização. Além de deixar mais evidente a polarização entre os partidos PP e PSOE, podemos notar também que existem diversos casos unicamente associados ao partido UM. Parte dessa divisão já era visível na rede dos envolvidos mostrada na Figura 2.1. Porém, a Figura 2.2 expõe ainda mais a separação entre os grupos associados aos indivíduos envolvidos em corrupção. Essa visualização

⁴O Apêndice A.4.1 apresenta brevemente o modelo de grafos aleatórios de Erdős–Rényi.

também destaca o papel importante dos partidos PP e PSOE, os quais são a grande maioria (102 dos 132) dos vértices da maior componente.

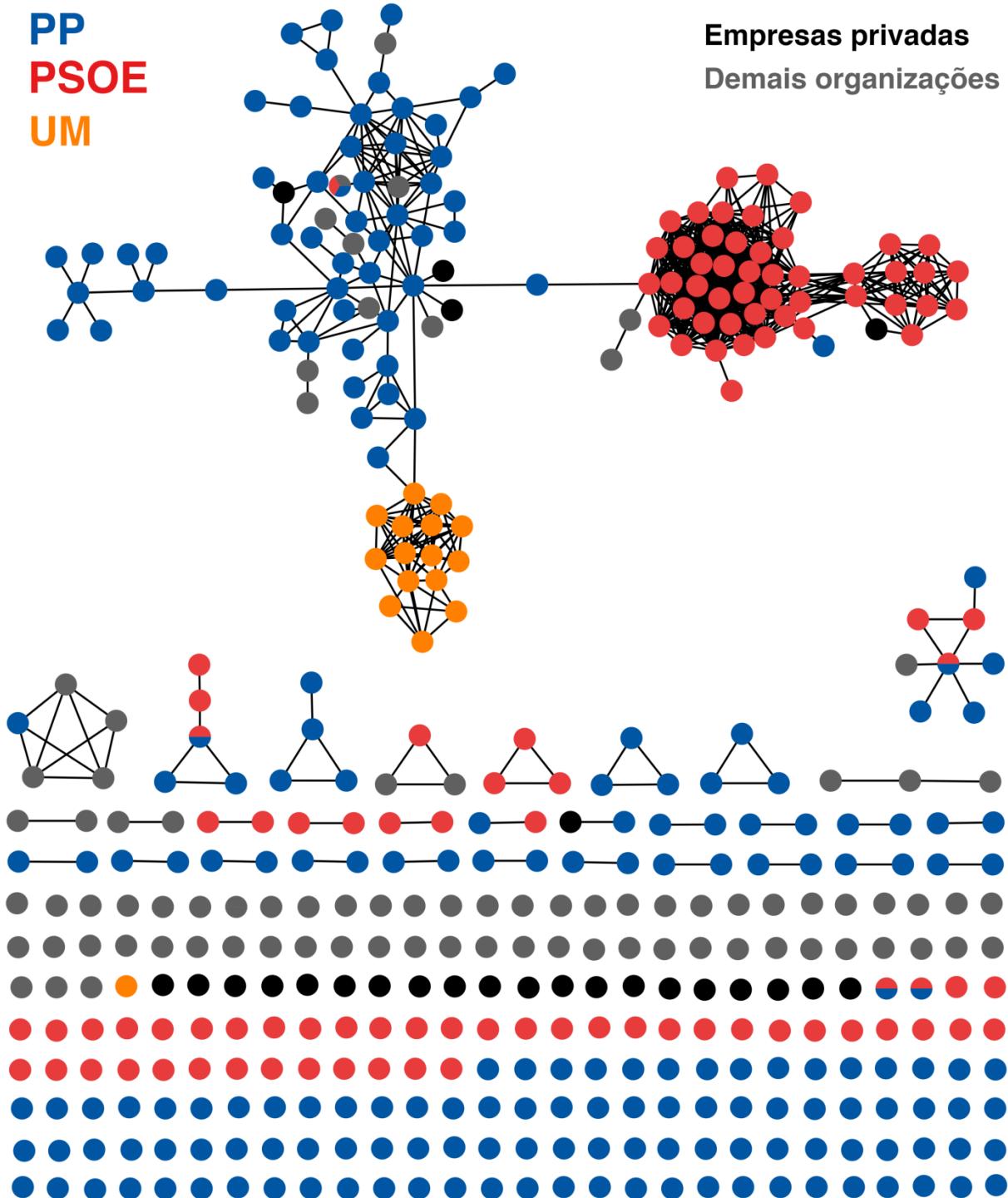


Figura 2.2: Visualização da rede dos casos de corrupção da Espanha. Nessa figura, cada vértice corresponde a um escândalo e dois casos de corrupção estão conectados se compartilham ao menos um envolvido. Essa rede é formada por 437 vértices e 750 ligações. As cores de cada vértice representam as organizações associadas a cada caso de corrupção. Casos com mais de uma organização são divididos igualmente nas cores de cada organização.

2.2 Vértices recorrentes e medidas de centralidade

Um aspecto interessante da rede de corrupção espanhola diz respeito ao comportamento de homens e mulheres envolvidos. Como já mencionamos, é bem documentado na literatura de criminologia a menor propensão feminina de envolvimento em práticas criminosas [29, 30]. De fato, a participação feminina na rede de corrupção espanhola é consideravelmente menor, uma vez que vértices mulheres correspondem apenas a 20% da rede. No entanto, a taxa de reincidência criminosa é apenas ligeiramente maior entre homens (9%) do que entre mulheres (7%). Aqui, a reincidência criminosa ocorre quando um vértice participa em mais de um caso de corrupção. Conforme discutiremos mais adiante, esses vértices são fundamentais para a estrutura da rede, uma vez que conectam diferentes partes da rede e funcionam como elos de comunicação. A Figura 2.3A mostra uma visualização da rede de corrupção espanhola com cores destacando os dois gêneros. Além disso, a Figura 2.3B mostra a evolução temporal do número de envolvidos em corrupção para cada gênero.

Para investigar padrões gerais relacionados ao gênero dos envolvidos, podemos calcular algumas medidas de centralidade. As Figuras 2.3C, 2.3D e 2.3E representam, respectivamente, os valores médios das centralidade de grau, intermediação e proximidade. Nesses gráficos, além de agrupar os resultados por gênero, diferenciamos também entre todos os vértices (recorrentes ou não) e apenas os vértices recorrentes. Cada valor médio possui uma barra de erro representando o erro padrão da média.

Os resultados indicam que indivíduos reincidentes de ambos os gêneros possuem maiores medidas de centralidade, apontando o papel central desses envolvidos para a formação e estrutura da rede. Essa diferença é menos acentuada para a centralidade de proximidade. Essa medida quantifica a existência de posições privilegiadas da rede, capazes de alcançar muitos vértices com menor distância. Assim, a diferença não muito expressiva dos valores médios desse parâmetro pode estar associada à polarização partidária da rede. Isto é, mesmo no caso de envolvidos reincidentes, essa polarização deve contribuir para reduzir o número de localizações privilegiadas na rede.

Os resultados das Figuras 2.3C, 2.3D e 2.3E também indicam que, apesar de serem minoria, as mulheres ocupam papéis semelhantes aos dos homens na rede de corrupção. Essas figuras mostram que os valores médios do número de conexões (Figura 2.3C), da centralidade de intermediação (Figura 2.3D) e da centralidade de proximidade (Figura 2.3E) são praticamente indistinguíveis entre homens e mulheres. Com exceção do valor médio do grau de todos os vértices, as barras de erro comparando homens e mulheres se sobrepõem. Para analisar as possíveis diferenças de gênero com maior rigor, usamos o teste de hipótese t de Welch⁵ para verificar a significância estatística das diferenças entre os valores médios das medidas de centralidade. A Tabela 2.2 mostra os p -valores desses testes.

⁵O Apêndice B.1.1 descreve esse teste em maior detalhe.

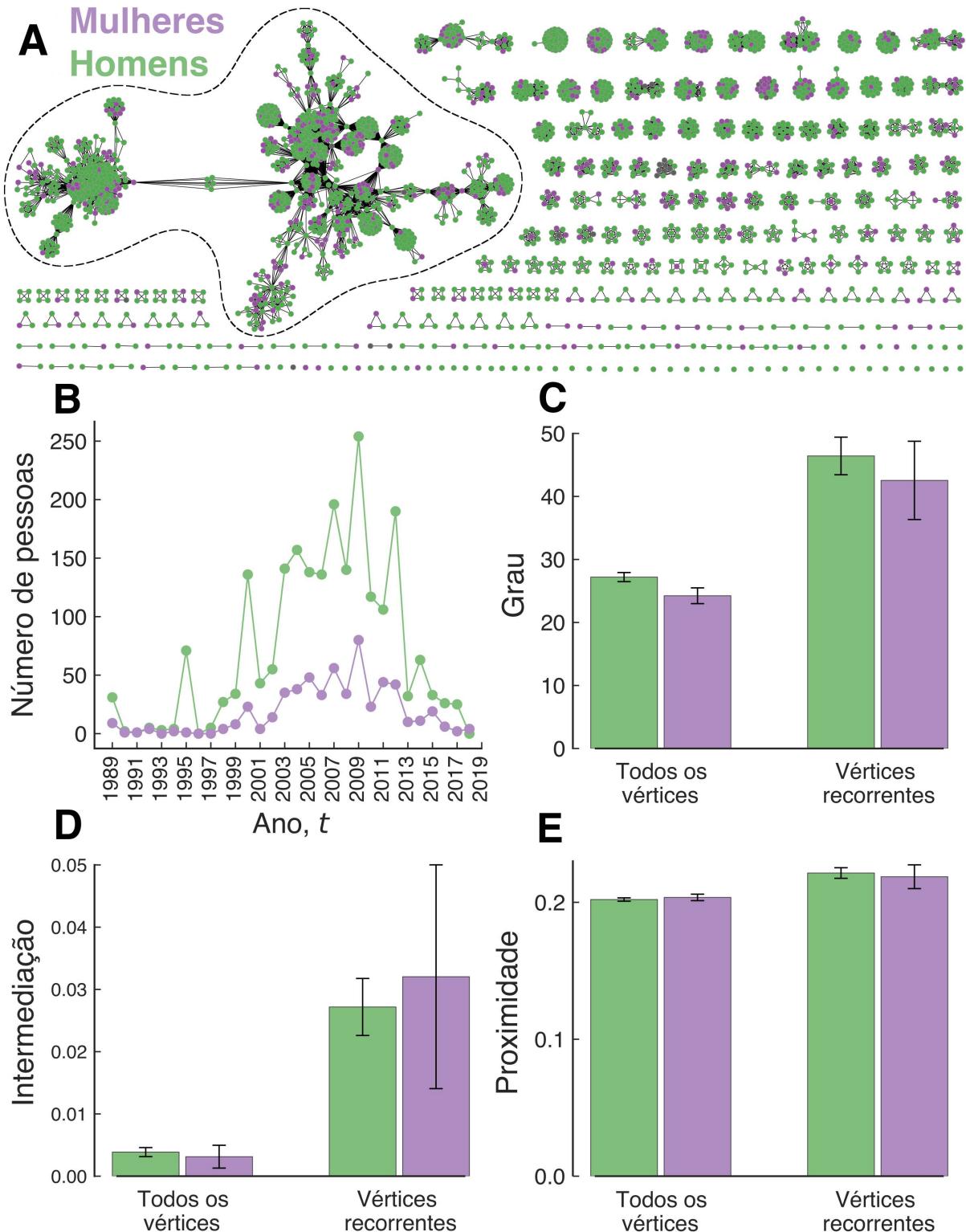


Figura 2.3: (A) Visualização da rede de corrupção espanhola destacando a participação de homens e mulheres. (B) Evolução temporal do número de homens e mulheres envolvidos nos casos de corrupção da Espanha. Valores médios das medidas de centralidade de (C) grau, (D) intermediação e (E) proximidade calculadas para a maior componente da rede. Cada painel é separado em todos os vértices (à esquerda) e vértices recorrentes (à direita) e as barras de erro correspondem ao erro padrão da média.

	Todos os vértices	Vértices recorrentes
Grau	0.041	0.577
Intermediação	0.709	0.796
Proximidade	0.563	0.777

Tabela 2.2: Os p -valores do teste t de Welch associados à comparação entre as médias de cada medida de centralidade para os dois gêneros.

Com exceção da comparação entre os valores médios do grau para todos os vértices, observamos que os p -valores são maiores do que 0.05. Entretanto, esse único p -valor menor do que 0.05 está muito próximo ao limiar de confiança de 95%. Nesse caso, não podemos concluir que existe uma diferença significante entre os valores médios das medidas de centralidade de homens e mulheres.

Outra informação relevante sobre os indivíduos da rede de corrupção se refere ao cargo de cada envolvido. Essa informação está disponível para mais de 95% dos indivíduos em nossos dados e pode ter relação com o papel do agente criminoso na rede. Para simplificar nossa análise, agrupamos os cargos dos indivíduos em 13 categorias: empresários, altos cargos privados, altos cargos públicos, funcionários públicos, vereadores, prefeitos, funcionários privados, assessores políticos, advogados, deputados, militares, senadores e outros. Esse agrupamento foi realizado manualmente. No caso de indivíduos que possuem mais de uma função ou competência, selecionamos a categoria de maior “influência”. Ademais, classificamos como “outros” aqueles envolvidos cuja ocupação é ambígua ou não se enquadra nas demais categorias.

O gráfico de barras da Figura 2.4A mostra o número total de pessoas por categoria de ocupação. Os círculos à direita são proporcionais ao número de vértices recorrentes que pertencem a cada grupo. Embora a categoria de empresários seja a mais numerosa, são os indivíduos vinculados a altos cargos públicos que possuem o maior número de reincidentes.

A Figura 2.4B mostra a quantidade total de pessoas em cada categoria ao considerar apenas a maior componente da rede de corrupção. Notamos que, mesmo considerando apenas os vértices da maior componente, os reincidentes da categoria altos cargos públicos ainda permanecem como o maior grupo. Ainda em relação aos envolvidos reincidentes, vale observar que assessores políticos, empresários e altos cargos públicos possuem grande parcela de seus vértices na maior componente (88%, 85% e 66%, respectivamente). Uma vez que envolvidos reincidentes são importantes para a associação entre dois ou mais escândalos, a grande fração desses indivíduos em poucas categorias indica a importância dessas categorias para a estrutura da rede de corrupção espanhola.

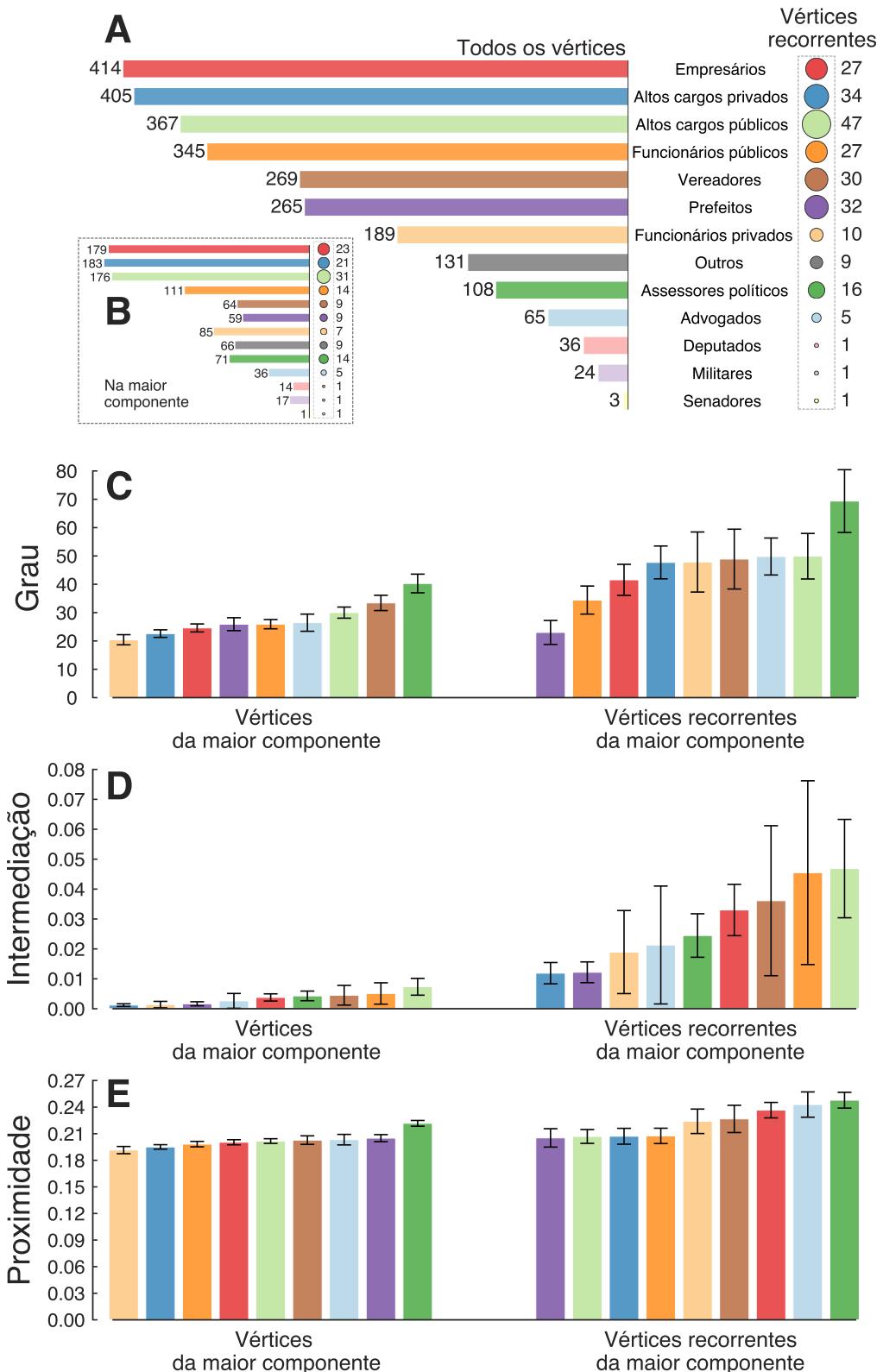


Figura 2.4: (A) Total de agentes envolvidos em corrupção agrupados por categoria de cargo. Os círculos à direita representam a quantidade de envolvidos reincidentes para cada categoria de cargo. (B) As mesmas quantidades do painel anterior ao considerar apenas os vértices da maior componente. Valores médios para as centralidades de (C) grau, (D) intermediação e (E) proximidade para os diferentes cargos, separados ainda entre todos os envolvidos da maior componente (à esquerda) e apenas reincidentes (à direita). As barras de erro representam o erro padrão da média.

Observamos ainda que os cargos de vereadores e prefeitos compõem uma parcela menor de vértices recorrentes (30% e 28%, respectivamente) e estão mais associados a escândalos menos conectados. De fato, é razoável assumir que a influência desses cargos seja mais localizada e restrita a escândalos de corrupção de suas regiões de atuação.

Com o objetivo de entender melhor o comportamento de cada categoria de ocupação na rede, estimamos algumas medidas de centralidade e agrupamos os resultados por categoria separando ainda os vértices recorrentes de todos os vértices. As Figuras 2.4C, 2.4D e 2.4E apresentam, respectivamente, os valores médios das centralidade de grau, intermediação e proximidade para os envolvidos de cada categoria. Para uma melhor visualização, selecionamos os 9 maiores valores de cada medida e separamos as figuras em duas partes. À esquerda, mostramos as médias para todos os vértices da maior componente e, à direita, apenas para os vértices recorrentes.

No quesito número de conexões, a Figura 2.4C revela os cargos mais centrais na rede de corrupção espanhola. Notamos que assessores políticos apresentam o maior grau médio ao considerar todos os vértices da maior componente e apenas os vértices recorrentes. No outro extremo, observamos que prefeitos possuem o menor número médio de conexões entre os vértices recorrentes. Apesar dessas diferenças importantes, o grau médio não varia muito entre os diferentes cargos. Vale notar também a tendência de vértices recorrentes apresentarem maiores valores médios para o grau em quase todos os casos mostrados na Figura 2.4C.

Em relação à centralidade de intermediação mostrada na Figura 2.4D, notamos que a maioria das categorias possui um erro padrão grande. Esse resultado indica uma dispersão considerável dos valores médios dessa medida e, portanto, a média da centralidade de intermediação pode não ser tão representativa nesses casos. Essa grande dispersão da média está associada à existência de vértices com grandes valores de intermediação (aqueles que conectam partes isoladas da rede) e a consequente emergência de diferentes ordens de grandeza para essa medida de centralidade. Apesar desse fato e contrariamente ao que ocorre para o grau, assessores políticos envolvidos em corrupção não apresentam os maiores valores de centralidade de intermediação. De fato, as três categorias que produzem os maiores valores médios de intermediação (e consequentemente conectam escândalos de maior relevância para a estrutura da rede) são altos cargos públicos, funcionários públicos e vereadores.

A medida de centralidade de proximidade mostrada na Figura 2.4E faz uma menor distinção entre os diferentes cargos dos agentes envolvidos em corrupção. Esse resultado indica novamente que a polarização partidária da rede impede uma maior variação da centralidade de proximidade.

No geral, a relação entre os cargos dos envolvidos e suas medidas de centralidades permite inferir algumas informações relevantes sobre a influência desses cargos na estrutura da rede. Por exemplo, o fato de assessores políticos apresentarem os maiores valores mé-

dios para as centralidades de grau e de proximidade, e valores relativamente menores para a centralidade de intermediação, indica que esses cargos fazem um menor número de conexões entre diferentes partes da rede. Assim, essas características sugerem que assessores políticos desempenham um papel mais local na rede de corrupção. Outra característica interessante se refere ao fato da categoria de funcionários públicos apresentar mais da metade dos envolvidos reincidentes na maior componente, um menor número de conexões em média (8^a posição) e um alto valor médio da centralidade de intermediação (2^a posição). Essas propriedades sugerem que, mesmo possuindo um número menor de conexões, essa categoria é responsável por conectar partes importantes da rede por meio de reincidência de práticas ilícitas. Por outro lado, os valores médios das centralidades obtidas para os vereadores apontam que, embora esse grupo disponha de poucos reincidentes na maior componente, eles apresentam grande centralidade de intermediação e de grau.

Não obstante, mesmo não apresentando valores médios altos para a centralidade de intermediação, empresários parecem desempenhar uma função relevante na rede, pois são vértices com grande parcela reincidentes na maior componente. Sendo assim, é possível que eles contribuam para a “comunicação” entre as partes internas da rede.

Além de ajudar a esclarecer o papel dos diferentes cargos para a estrutura da rede, essas informações podem expor aspectos vulneráveis desse sistema. Com a finalidade de entender melhor as fragilidades de redes de corrupção e, possivelmente, embasar medidas que impeçam seu funcionamento, a próxima seção faz uso de nossos achados empíricos sobre os cargos e medidas de centralidade para criar ataques direcionados visando o desmantelamento mais efetivo da rede de corrupção.

2.3 Ataques direcionados por cargo e desmantelamento da rede

Uma das questões relevantes sobre a robustez de redes complexas se refere ao quanto essas redes são resilientes à remoção de vértices ou ligações. Encontrar o menor conjunto de vértices que removidos fragmentam a rede com maior eficácia é importante porque auxilia na identificação de vulnerabilidades, permite ataques mais efetivos e contribui para a construção de redes mais robustas. A eficácia na fragmentação de uma rede pode ser medida pelo tamanho da maior componente ao passo que as ligações são removidas. Diversos esforços da comunidade científica têm sido destinados a esse propósito, incluindo a análise de vários tipos de redes, abordagens [34–36] e, até mesmo, a recente proposição de um método geral e eficiente para o desmantelamento de redes [37].

Em relação à uma rede de crimes federais, Cunha *et al.* [21] compararam diversas abordagens de fragmentação e verificaram que esse tipo de sistema é bastante suscetível à remoção de vértices que conectam diferentes comunidades (ou módulos). No trabalho,

os autores argumentam que agentes conectando diferentes comunidades provavelmente possuem cargos específicos (advogados, contadores e outros). Além disso, esses vértices apresentam medidas de centralidade de intermediação mais altas e, portanto, atuam como elos entre diferentes partes da rede. Destacamos também um outro trabalho que analisa a resiliência de redes criminosas explorando o processo de fragmentação de uma rede mafiosa italiana, conhecida como *Cosa nostra* [38]. Essa análise mostra que a remoção de apenas 5% dos vértices com os maiores valores de centralidade de intermediação já é suficiente para reduzir o número de conexões em 70%.

De forma geral, os dois trabalhos citados anteriormente possuem em comum o fato de que a medida de centralidade de intermediação se mostrou bastante efetiva para o desmantelamento das redes criminosas. Contudo, a questão de como atacar eficientemente essas redes criminosas ainda não está completamente resolvida. Além disso, nem sempre a remoção, prisão, ou inibição de certos envolvidos é viável ou possível. Naturalmente, o custo para direcionar ataques aos envolvidos mais importantes da rede é maior e costuma requerer grandes ações por parte das autoridades policiais. Ciente desses aspectos, propomos utilizar as informações sobre o cargo de cada envolvido para realizar ataques na rede de corrupção espanhola.

A Figura 2.5 mostra a fração da maior componente em função do número de vértices removidos de acordo com uma classificação baseada em três medidas de centralidade (grau, intermediação e proximidade). Para cada medida de centralidade, consideramos ainda duas estratégias: remoção a partir de todos os vértices (independente se são recorrentes ou não) e remoção apenas de vértices recorrentes. Para facilitar a comparação entre esses dois procedimentos, dividimos cada gráfico em duas partes de acordo com cada estratégia de desmantelamento.

Selecionamos três categorias (funcionários públicos de alto escalão, funcionários públicos e empresários) as quais verificamos serem mais eficientes para desmantelar a rede. Mais especificamente, para uma dada categoria, a remoção dos vértices é realizada em ordem decrescente de valores da medida de centralidade. Para ilustrar esses valores, cada painel contém um código de cores associado a cada categoria. Esse código é usado nas curvas da Figura 2.5 e corresponde à variação nos valores de cada medida de centralidade.

Para cada uma das estratégias, estimamos também o valor médio (em 100 realizações) da fração da maior componente ao remover os vértices de maneira aleatória (curvas em preto). Essa última quantidade serve de linha de base para a comparação com outras abordagens baseadas nos cargos e nas medidas de centralidade. Acompanhando as curvas relacionadas à remoção aleatória, calculamos uma banda de confiança que representa o erro padrão dos valores médios (regiões em cinza). Esperamos que quanto mais eficaz for a abordagem de remoção, mais rapidamente a rede deve ser fragmentada quando comparada ao caso aleatório.

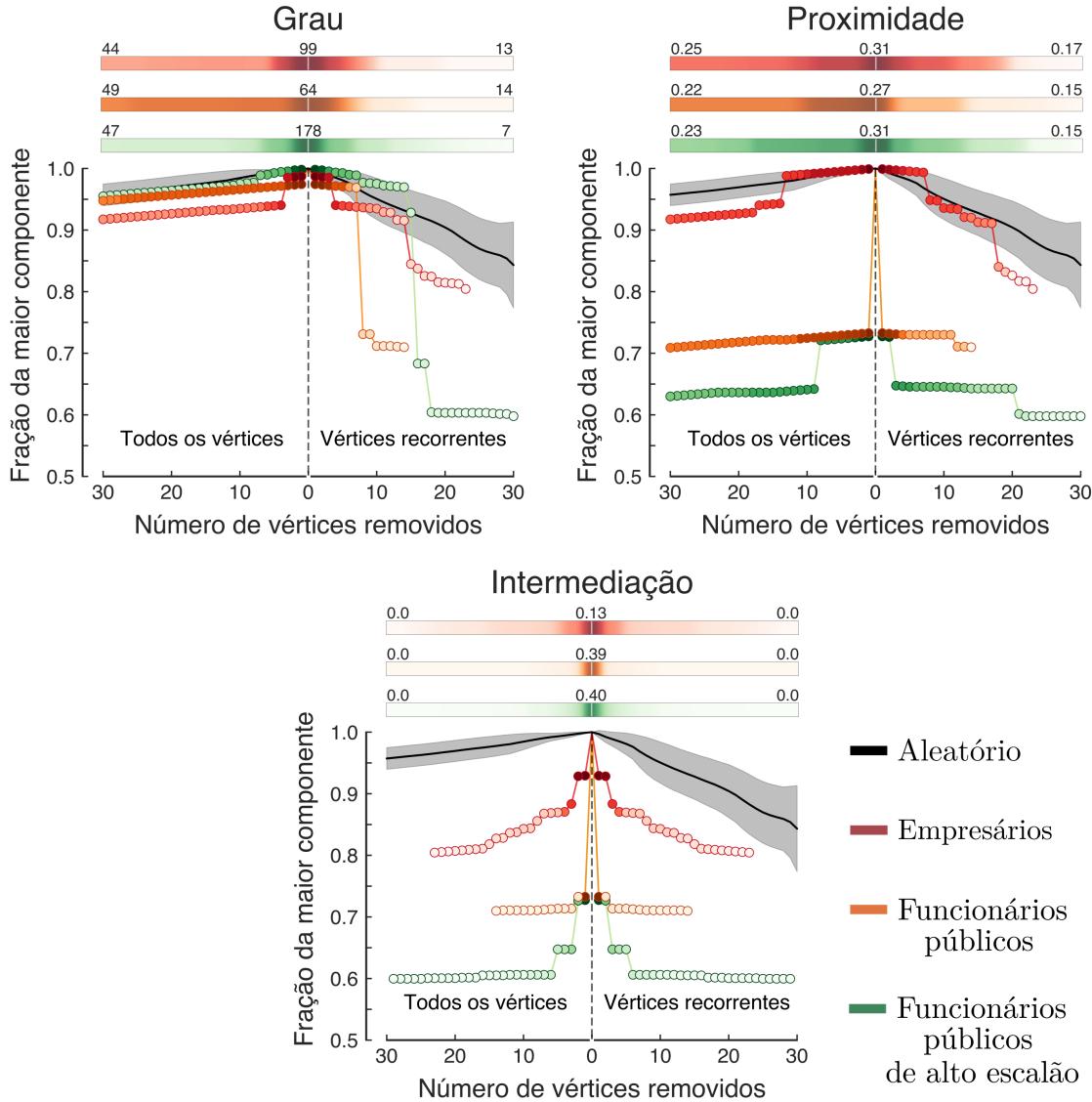


Figura 2.5: Desmantelamento da rede de corrupção espanhola ao remover os envolvidos de três categorias de cargos em ordem decrescente de três medidas de centralidade (grau, proximidade e intermediação). Cada painel é composto por duas partes: à esquerda, apresentamos a fração da maior componente em função do número de vértices removidos ao considerar todos os vértices; à direita, mostramos a mesma quantidade estimada usando apenas os vértices recorrentes. As categorias escolhidas são as que fragmentam a rede mais rapidamente. As barras coloridas e os códigos de cores das curvas correspondem aos valores de centralidade dos vértices removidos. As curvas em preto representam os valores médios das frações obtidas ao remover os vértices de maneira aleatória e as regiões em cinza representam o erro padrão desses valores médios.

Algumas considerações sobre a Figura 2.5 merecem destaque. Comparando ambas as estratégias, notamos que o número máximo de vértices removidos para um dado cargo pode variar. Isso ocorre porque removemos apenas os vértices de cada categoria com medida de centralidade maior do que zero. Por exemplo, na remoção de vértices recorrentes por ordem decrescente de grau, embora existam 30 envolvidos na categoria de funcionários

públicos de alto escalão, a categoria de funcionários públicos apresenta apenas 14 envolvidos que satisfazem o critério anterior. Além disso, notamos que as curvas são simétricas em torno do eixo central no caso da medida de intermediação. Nesse caso, os envolvidos que possuem intermediação maior do que zero também são reincidentes⁶ e, portanto, não existe diferença entre as duas estratégias.

Comparando o desmantelamento entre as três medidas de centralidade, observamos que todos os cenários possuem uma eficácia maior quando vértices recorrentes são removidos. Mais especificamente, o desmantelamento baseado na centralidade de intermediação apresenta a maior eficiência: a remoção de apenas 5 funcionários públicos de alto escalão é suficiente para reduzir a maior componente da rede a 60% do seu tamanho original. Em relação aos funcionários públicos, a remoção de 3 vértices já é suficiente para reduzir o tamanho da maior componente em quase 30%.

Por outro lado, a abordagem menos eficiente de fragmentação da rede é a baseada no grau do vértice. A remoção de vértices com grande número de conexões não afeta substancialmente o tamanho da maior componente. A informação de recorrência dos vértices contribui para uma maior eficácia dessa abordagem. Entretanto, mesmo usando essa informação, necessitamos de um número maior de vértices para obter porcentagens de fragmentação similares às obtidas para outras medidas de centralidade.

No meio termo temos o desmantelamento baseado na centralidade de proximidade. Para envolvidos reincidentes que pertencem à categoria de funcionários públicos de alto escalão e funcionários públicos, a remoção de vértices fragmenta a rede de maneira similar à centralidade de intermediação. Já em relação aos cargos de empresários, esse processo aparenta ser menos eficaz e similar à remoção de empresários baseada na centralidade de grau.

Assim, em ordem decrescente de eficácia, as categorias de funcionários públicos de alto escalão, funcionários públicos e empresários, apresentam os envolvidos reincidentes com maior potencial de dano para a estrutura da rede de corrupção espanhola. Considerando todos esses aspectos, podemos concluir que a reincidência criminosa dos envolvidos apresenta um papel crucial para a estrutura e vulnerabilidade da rede de corrupção da Espanha. Essa análise também sugere que o monitoramento de indivíduos reincidentes (especialmente aqueles que ocupam os três cargos listados anteriormente) pode ser muito útil para ampliar o conhecimento das associações criminosas e, consequentemente, reduzir a prática de corrupção política.

⁶Embora todos os envolvidos com centralidade de intermediação maior do que zero sejam reincidentes, nem todos os reincidentes possuem intermediação maior do que zero (apenas dois indivíduos em toda a rede).

Capítulo 3

Dinâmica de crescimento da rede de corrupção espanhola

O tratamento de redes complexas como sistemas complexos dinâmicos permite compreender os principais mecanismos responsáveis pela formação e desenvolvimento de suas propriedades. Porém, o estudo realizado até agora tratou a rede de corrupção da Espanha de forma estática e considerou apenas seu estágio final. Essa estrutura final é resultado de um processo de crescimento ao longo dos 30 anos cobertos por nossa base de dados. Os vértices e escândalos associados à rede foram progressivamente descobertos e incorporados à rede e a investigação desse processo pode revelar padrões importantes sobre como operam essas redes criminosas. Neste capítulo, estudamos a rede de corrupção espanhola considerando seu aspecto temporal.

3.1 Visualizando a dinâmica da rede

Para dar início à nossa análise, a Figura 3.1 apresenta uma visualização do crescimento da rede de corrupção espanhola. Essa figura mostra a rede para quatro anos de sua evolução: 1994, 2002, 2010 e 2018. Observa-se que a rede se desenvolve devido ao surgimento de novos escândalos e de envolvidos reincidentes que conectam esses escândalos. Assim, a rede desenvolve uma estrutura com a emergência de algumas grandes componentes conectadas e várias outras componentes menores.

Além dessa visualização, criamos um vídeo¹ mostrando os 30 anos do crescimento da rede de corrupção espanhola. Esse vídeo realça o aspecto não trivial da evolução dessa rede. Portanto, para extrair padrões e informações sobre a dinâmica dessa rede, é necessário uma análise mais detalhada do seu processo de evolução temporal.

¹<https://www.youtube.com/watch?v=AgmoKnTy25k>

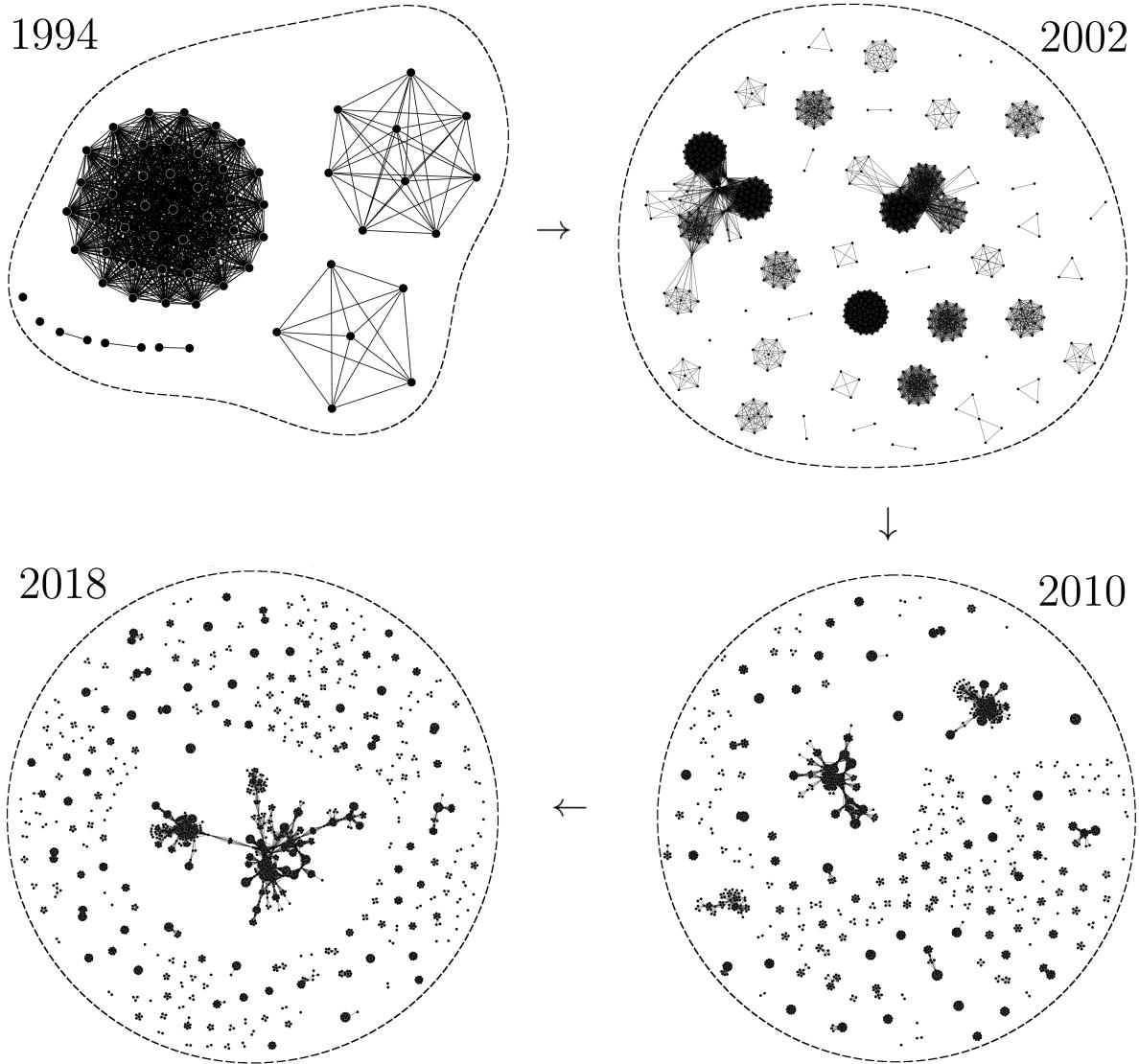


Figura 3.1: Visualização do aspecto temporal da rede de corrupção espanhola. Mostramos a evolução da rede para os anos de 1994, 2002, 2010 e 2018. O crescimento dessa rede se deve ao surgimento de novos escândalos de corrupção e de envolvidos reincidentes que conectam esses casos.

3.2 Evolução temporal da distribuição de grau

Redes de tráfico de drogas [39], terrorismo [40] e, mais recentemente, corrupção brasileira [18], possuem em comum distribuições de grau aproximadas pela distribuição exponencial. Para verificar se a rede de corrupção espanhola também apresenta esse padrão, estimamos sua distribuição ao longo dos anos. Em escala mono-logarítmica, a Figura 3.2A mostra essa distribuição acumulada do grau dos vértices da rede para os anos de 1995, 2001 e 2018. Nessa figura, as linhas tracejadas representam distribuições exponenciais ajustadas aos dados via método de máxima verossimilhança. Além disso, os valores do grau característico estão indicados ao lado de cada curva ajustada. Embora o acordo

não seja perfeito, notamos que a distribuição exponencial aproxima razoavelmente bem o comportamento das distribuições empíricas para os anos analisados.

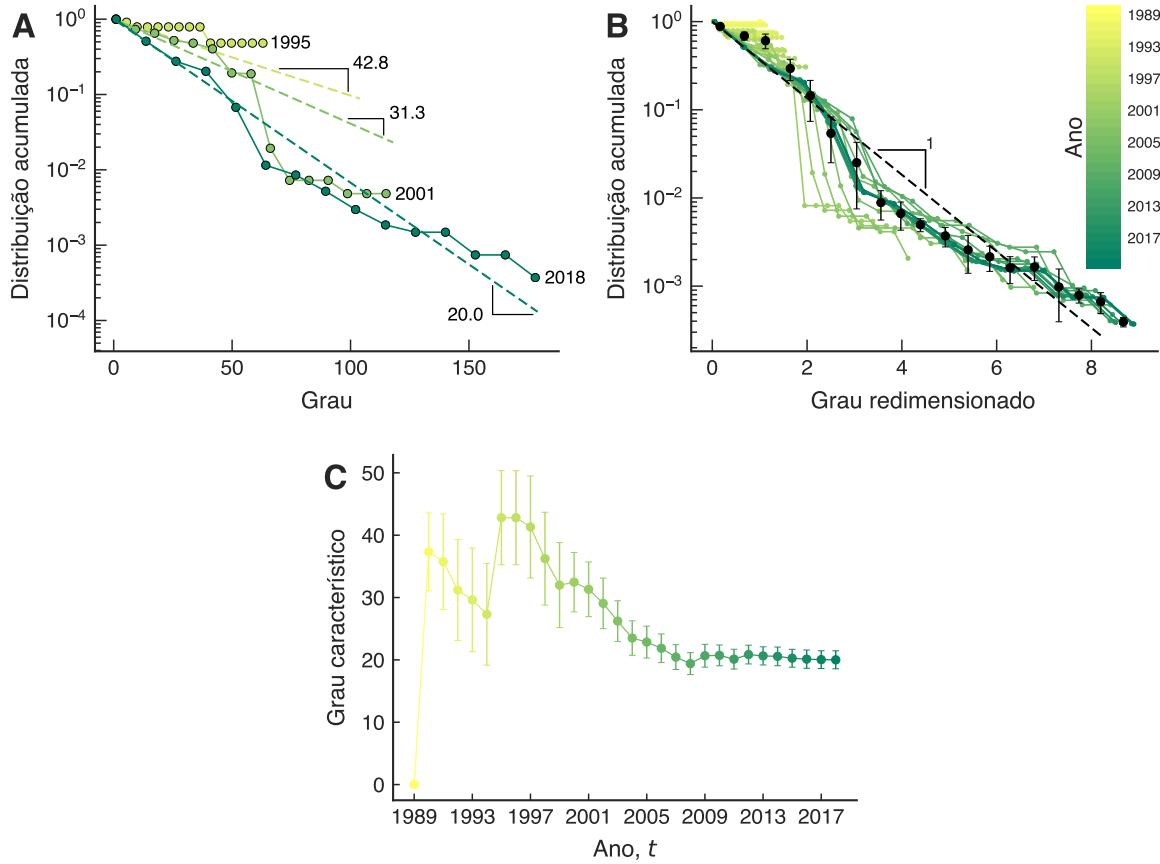


Figura 3.2: (A) Distribuição acumulada do grau para três diferentes estágios da rede de corrupção espanhola. (B) Distribuição acumulada do grau reescalado pelo valor médio do grau da rede em cada ano. Para fins de comparação, a linha preta tracejada representa a distribuição exponencial com grau característico unitário e os pontos em preto representam valores médios sobre todas as distribuições. Além disso, a cor de cada curva corresponde ao ano de cada rede, conforme o código de cores presente na barra à direita. (C) Série temporal do grau característico ao longo dos 30 anos de evolução da rede de corrupção da Espanha. As barras de erro indicam intervalos de confiança de 95% estimados por *bootstrap*.

Ainda nesse contexto, é interessante saber se a distribuição de grau para os demais anos apresenta uma tendência exponencial similar a observada na Figura 3.2A. Para analisar essa questão, calculamos as distribuições acumuladas do grau para cada ano da rede após reescalar o grau dos vértices pelo valor médio do grau nos anos correspondentes. Ou seja, para cada ano, os graus dos vértices são divididos por seu respectivo valor médio. Considerando a hipótese exponencial², essa operação de reescala deve fazer as distribuições de grau dos diferentes anos colapsarem em uma única distribuição com grau caracterís-

²O Apêndice B.3 apresenta detalhes sobre a distribuição exponencial.

tico unitário. A Figura 3.2B mostra essas distribuições de grau reescalado para todos os anos da rede. Notamos um bom colapso das distribuições, indicando que a distribuição exponencial é uma boa aproximação para os dados. Para melhor observar esse comportamento, calculamos também o valor médio dessas distribuições, conforme indicado pelos pontos pretos na Figura 3.2B. Considerando que as barras de erro compõem um intervalo de confiança de 95% obtido por *bootstrap*³, podemos constatar um acordo razoável entre o comportamento médio das distribuições e a distribuição exponencial.

A Figura 3.2C mostra a evolução temporal do grau característico ajustado para cada ano. Sob a hipótese exponencial, é importante saber se o grau característico exibe alguma tendência dinâmica. Diferentemente do comportamento da rede de corrupção brasileira [18], o grau característico no caso espanhol não apresenta transições entre platôs bem definidos. Entretanto, podemos identificar alguns estágios diferentes durante a evolução temporal desse parâmetro. Entre 1990 e 1994, observamos que o grau característico apresenta uma tendência decrescente ao redor do valor 30. Em 1995, o grau característico sofre um aumento abrupto, seguido por uma tendência decrescente que persiste até por volta do ano 2010. Após esse ano, o grau característico fica estável em um valor ligeiramente superior a 20.

Curiosamente, o grau característico da rede espanhola estabilizou em um valor bem próximo ao reportado para a rede de corrupção brasileira [18]. No caso brasileiro, o grau característico sofre algumas mudanças abruptas até estabilizar em ≈ 18 . Essa semelhança entre as duas redes de corrupção pode ser um indicativo de um comportamento universal dessas redes, isto é, apesar das especificidades de cada país e de seus casos de corrupção, a forma exponencial da distribuição de grau e o seu grau característico não mudam muito entre os dois sistemas. Isso indica uma possível semelhança nas redes de corrupção de ambos os países, de forma que seus padrões de conexão parecem independente de aspectos específicos de cada país.

3.3 Dinâmica das grandes componentes da rede

Em muitos casos, a evolução de redes complexas leva à emergência de grandes componentes conectadas. Essa emergência pode ocorrer por meio de um processo parecido com uma transições de fase [22], no qual componentes da rede se agrupam para criar uma componente gigante.

Essas transições ocorrem devido ao surgimento de novas ligações. Nesse cenário, uma maneira de observar esse fenômeno de coalescência é por meio da evolução temporal do tamanho (número de vértices) das maiores componentes da rede. No nosso caso, é importante observar que redes criminosas evoluem no tempo devido ao surgimento e descoberta de novas atividades criminosas. Isso faz com que a rede tenha propriedades

³O Apêndice B.4 descreve o procedimento de *bootstrap* com maior detalhe.

dinâmicas importantes para a compreensão dos processos do crime organizado. No caso da corrupção política brasileira, Ribeiro *et al.* [18] mostram que a evolução da rede ocorre por meio de processos de coalescência de componentes isoladas da rede, nas quais existe um número muito reduzido de indivíduos responsáveis pela interligação de diferentes grupos de escândalos de corrupção.

Conforme ilustra a Figura 3.3, nossas investigações sobre a rede de corrupção da Espanha também apontam para a presença de um fenômeno tipo coalescência no crescimento dessa rede. Mais especificamente, essa rede passou por uma transição que resultou na coalescência das três maiores componentes presentes no ano de 2011, formando a componente gigante de 2012. A Figura 3.3A apresenta a evolução temporal do tamanho das duas maiores componentes da rede durante os 30 anos de nossa base de dados. Podemos observar que, entre 2011 e 2012, o tamanho da maior componente aumenta substancialmente, ao mesmo tempo que a segunda maior componente diminui em tamanho. Considerando que os envolvidos nunca são removidos da rede, esse gráfico indica a junção das duas maiores componentes em uma só. Um comportamento muito similar a esse é observado em transições de fase de modelos de percolação [41].

Para investigar melhor esse comportamento, a Figura 3.3B mostra a variação anual do tamanho das duas maiores componentes. Para o ano de 2012, observamos um valor máximo para a maior componente e, concomitantemente, um valor mínimo para a segunda maior componente. Esse gráfico identifica precisamente a transição abrupta na rede de corrupção espanhola, isto é, a maior componente aumenta bruscamente de tamanho porque alguns escândalos fazem as duas maiores componentes de 2011 se juntarem.

Uma terceira análise mais qualitativa pode ser feita observando as componentes graficamente. Para isso, identificamos as maiores componentes da rede para o ano de 2011 e descobrimos os escândalos e envolvidos que as conectam em 2012. De fato, como mostra a Figura 3.3C, é possível confirmar que as duas maiores componentes da rede de 2011 (primeira em roxo e segunda em verde) se conectam para formar a componente gigante em 2012. Além disso, essa figura revela que a terceira maior componente da rede (em laranja) também participou desse processo.

Para destacar os principais vértices responsáveis por esse fenômeno, as três maiores componentes de 2011 mostradas na Figura 3.3C possuem alguns vértices em cores diferentes. Em especial, a componente gigante contém 6 vértices amarelos e um único vértice azul. A segunda maior componente possui um vértice central em vermelho e a terceira maior componente possui 2 vértices de cores diferentes (azul e amarelo). Esses poucos envolvidos destacados se tornam reincidentes nos escândalos de 2012 e, portanto, são os responsáveis pela formação da componente gigante desse ano.

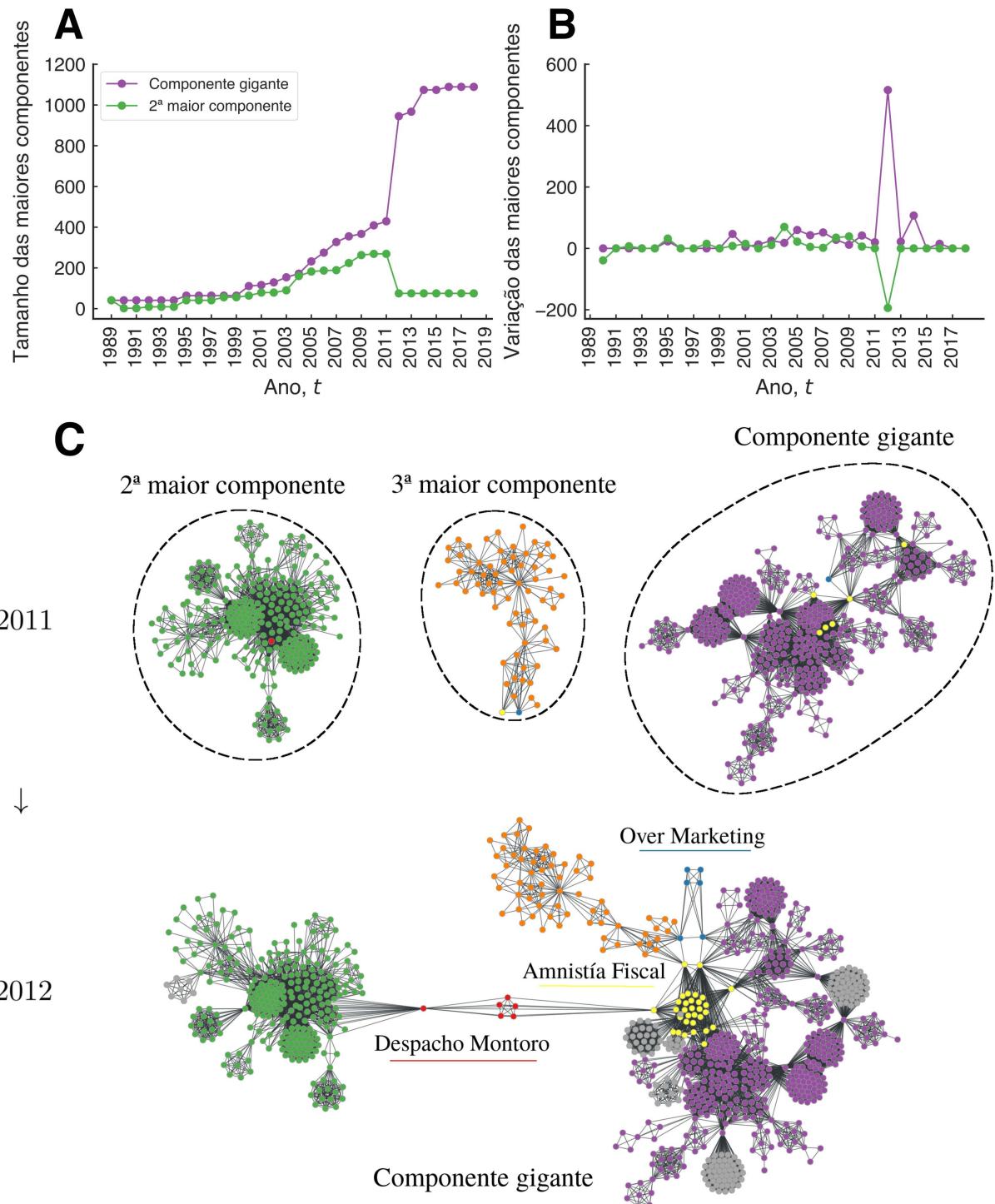


Figura 3.3: (A) Evolução temporal do tamanho das duas maiores componentes da rede de corrupção espanhola. (B) Variação anual do tamanho dessas duas componentes. (C) Processo de coalescência entre as maiores componentes da rede. Observamos o surgimento de três novos casos de corrupção em 2012: *Despacho Montoro*, *Amnistía Fiscal* e *Over marketing*. Os envolvidos desses três escândalos estão coloridos de vermelho, amarelo e azul, respectivamente. Alguns poucos agentes presentes na rede em 2011 também estão implicados nesses casos. Isso faz com que ocorra a coalescência entre as três maiores componentes da rede do ano de 2011, as quais passam a compor a maior componente do ano de 2012. Nessa figura, as cores roxa, verde e laranja representam respectivamente a primeira, segunda e terceira maior componente da rede. Os demais novos envolvidos em 2012 estão coloridos de cinza.

A coalescência resulta do surgimento de três novos casos de corrupção (*Despacho Montoro*, *Amnistía Fiscal* e *Over Marketing*) em 2012. A componente gigante desse ano possui os vértices desses escândalos nas cores vermelho, amarelo e azul, respectivamente. Já os vértices em cinza representam novos implicados não envolvidos em escândalos anteriores.

Com esses vértices e casos em destaque, notamos que o escândalo de corrupção conhecido como *Despacho Montoro* foi responsável por conectar as duas maiores componentes de 2011. Nesse contexto, um implicado presente na segunda maior componente do ano de 2011 se tornou reincidente ao se envolver no caso *Despacho Montoro* de 2012. Esse caso, por sua vez, se conectou à segunda maior componente do ano de 2011 por meio de outro envolvido reincidente presente no escândalo *Amnistía Fiscal*. Por fim, a junção da terceira com a primeira maior componente foi causada por envolvidos reincidentes presentes nos escândalos *Amnistía Fiscal* e *Over Marketing*.

Para contextualizar, apresentamos uma breve descrição sobre esses três escândalos:

- O caso de corrupção conhecido como *Despacho Montoro* [42] ocorreu em 2012 e revelou atividades ilícitas associadas a membros de uma empresa de consultoria fundada em 2006 pelo então Ministro da Fazenda da Espanha. Dentre as supostas atividades criminosas, se destacam prevaricação administrativa, desvios de fundos públicos, tráfico de influência e falsificação de documentos.
- O caso *Amnistía Fiscal* [43] se refere à anistia fiscal aprovada pelo governo do Partido Popular e que depois foi declarada inconstitucional. Essa medida de 2012 foi lançada pelo então Ministro da Fazenda espanhol e, durante o tempo em que permaneceu válida, foi utilizada por indivíduos como uma maneira de lavagem de dinheiro.
- O escândalo de corrupção conhecido como *Over Marketing* [44], também ocorrido em 2012, envolveu um financiamento ilegal de despesas eleitorais do Partido Popular de Baleares. Esse financiamento foi realizado por meio de empresas locais e contratos públicos.

Para apresentar mais informações dos escândalos relacionadas à coalescência observada na Figura 3.3, podemos construir uma rede similar a da Figura 2.2. Essa representação é útil porque evidencia quais escândalos se conectam e mostra seus respectivos partidos. Sendo assim, a Figura 3.4 exibe a coalescência por meio da rede dos escândalos de corrupção espanhóis. Podemos observar que cada componente é composta majoritariamente por um dos seguintes partidos: PSOE (vermelho), PP (azul) ou UM (laranja). Portanto, a coalescência da rede produziu uma componente gigante com esses três partidos. Nessa figura, os poucos vértices associados a outras organizações estão coloridos de cinza.

Considerando essas informações, podemos concluir que a transição observada na rede de corrupção espanhola pode ser explicada pela conexão entre suas componentes polarizadas do ponto de vista partidário. Cada uma dessas componentes da rede cresceu de forma

independente ao longo dos anos e apresenta envolvidos e escândalos majoritariamente associados a uma única organização. Alguns poucos escândalos e envolvidos de 2012 fizeram essas partes se unirem, causando a transição que observamos na Figura 3.3.

Assim, embora vários escândalos tenham contribuído para o crescimento da rede, há um processo específico que se destaca. Esse processo representa o surgimento de três escândalos de corrupção que conectam envolvidos das três componentes da rede associadas aos grandes partidos políticos espanhóis. Diferente do processo de crescimento da rede de corrupção brasileira, na qual observa-se ao menos três variações repentina no tamanho da maior componente [18], a evolução da rede de corrupção espanhola apresenta uma única transição abrupta.

Esse resultado indica que a polarização partidária da rede de corrupção da Espanha desempenhou um papel crucial no crescimento da rede. Até 2011, as componentes eram compostas em sua maioria por um único partido e crescia separadamente. Ao se juntarem em 2012, esse processo gerou uma mudança substancial na rede.

Em linhas gerais, podemos notar que existe uma semelhança na coalescência das redes de corrupção brasileira e espanhola. O processo de coalescência, mesmo que possivelmente motivado por diferentes circunstâncias, depende de um pequeno número de envolvidos reincidientes. Essa aparente universalidade indica que o modo de operação de organizações ligadas à corrupção política é similar entre nações e contextos políticos diferentes.

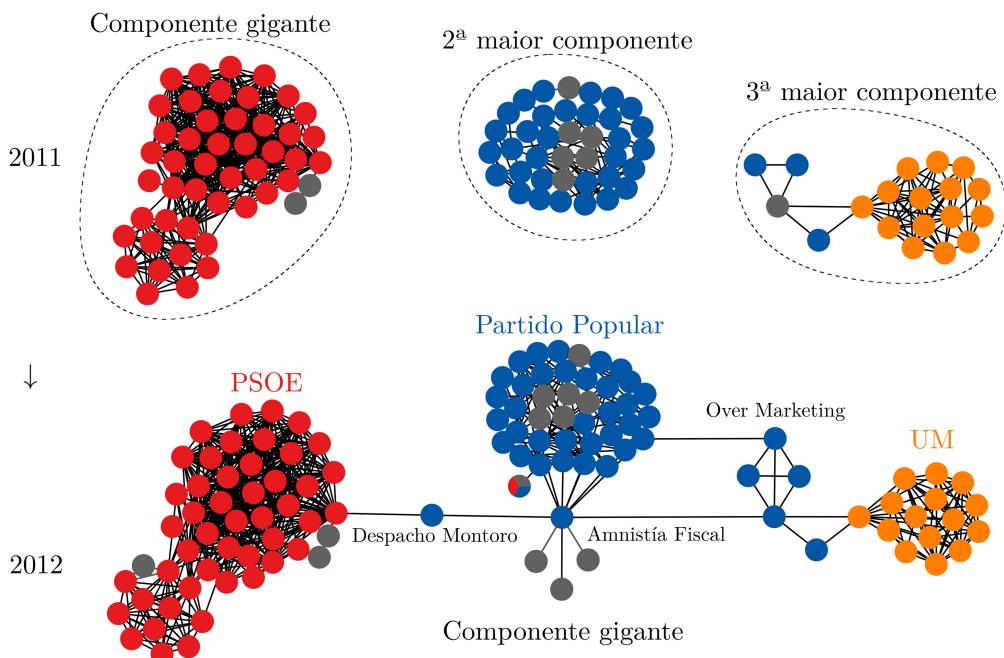


Figura 3.4: Visualização da coalescência observada na rede dos casos de corrupção da Espanha. Nessa figura, os partidos PSOE (vermelho), PP (azul) e UM (laranja) compõem majoritariamente uma das três componentes presentes em 2011. O processo de coalescência é causado por três novos escândalos de 2012 e resulta em uma componente gigante composta majoritariamente por esses três partidos. Nessa figura, os vértices em cinza representam casos associados às demais organizações presentes em nossa base de dados.

3.4 Estrutura de comunidades

A estrutura de comunidades (ou módulos) é uma propriedade encontrada em diversos tipos de redes [45]. De forma simplificada, esses módulos ou comunidades podem ser pensados como grupos de vértices mais densamente conectados entre si do que com outros grupos de vértices da rede [46]. A detecção dessas estruturas pode fornecer informações importantes sobre a rede. Por exemplo, estruturas de comunidades podem ser levadas em conta para melhorar a predição de novas ligações [47] e estudar processos de propagação de epidemias [48] e informações [49].

Em relação às redes criminosas, Calderoni *et al.* [50] argumentam que organizações criminosas possuem a tendência de formar comunidades com objetivo de reduzir o risco de vazamento de informações. No caso de redes criminosas brasileiras, Cunha *et al.* observaram que redes de crimes federais possuem padrões modulares [21] e que informações sobre esses módulos são úteis para ataques direcionados na rede [36]. Por sua vez, Ribeiro *et al.* mostraram que agentes envolvidos na rede de corrupção brasileira também formam estruturas modulares [18]. Em particular, Ribeiro *et al.* sugerem que o fato do número de escândalos (65) ser aproximadamente duas vezes o número de módulos (27), indica que existem vários escândalos que poderiam ser considerados como um só. No caso da rede de corrupção espanhola, sabemos que alguns escândalos possuem vários desdobramentos (por exemplo, os casos ERE e Gürtel). Consequentemente, podemos imaginar que a estrutura de comunidades dessa rede também apresenta, em média, mais de um escândalo por comunidade.

Com o objetivo de encontrar a estrutura de comunidades da rede de corrupção espanhola, utilizamos a abordagem bayesiana de blocos estocásticos⁴ desenvolvida por Peixoto [51]. Além de ser não paramétrico, esse modelo permite encontrar um grande número de comunidades. Mais especificamente, consideramos uma extensão dessa abordagem conhecida como modelo de blocos estocásticos aninhados. Dessa forma, é possível encontrar uma hierarquia de comunidades: comunidades dentro de comunidades. Para aplicar essa abordagem aos nossos dados, usamos a biblioteca *graph-tool* [51], que implementa esses métodos para a linguagem de programação *Python*.

A Figura 3.5A apresenta uma visualização da maior componente da rede de corrupção espanhola com destaque para sua estrutura de comunidades. As cores representam diferentes comunidades e os vértices que pertencem à uma mesma comunidade estão posicionados mais próximos uns dos outros. Nessa figura, podemos notar que existem certos agrupamentos de ligações (padrão gerado por um algoritmo conhecido como *edge-bundling* [52]). Isso ocorre quando a quantidade de ligações é grande e essas ligações estão bastante “próximas”. Além disso, um olhar mais atento revela que entre as comunidades

⁴O Apêndice A.5 descreve brevemente essa abordagem, a qual iremos nos referir por SBM (do inglês, *Stochastic Block Model*).

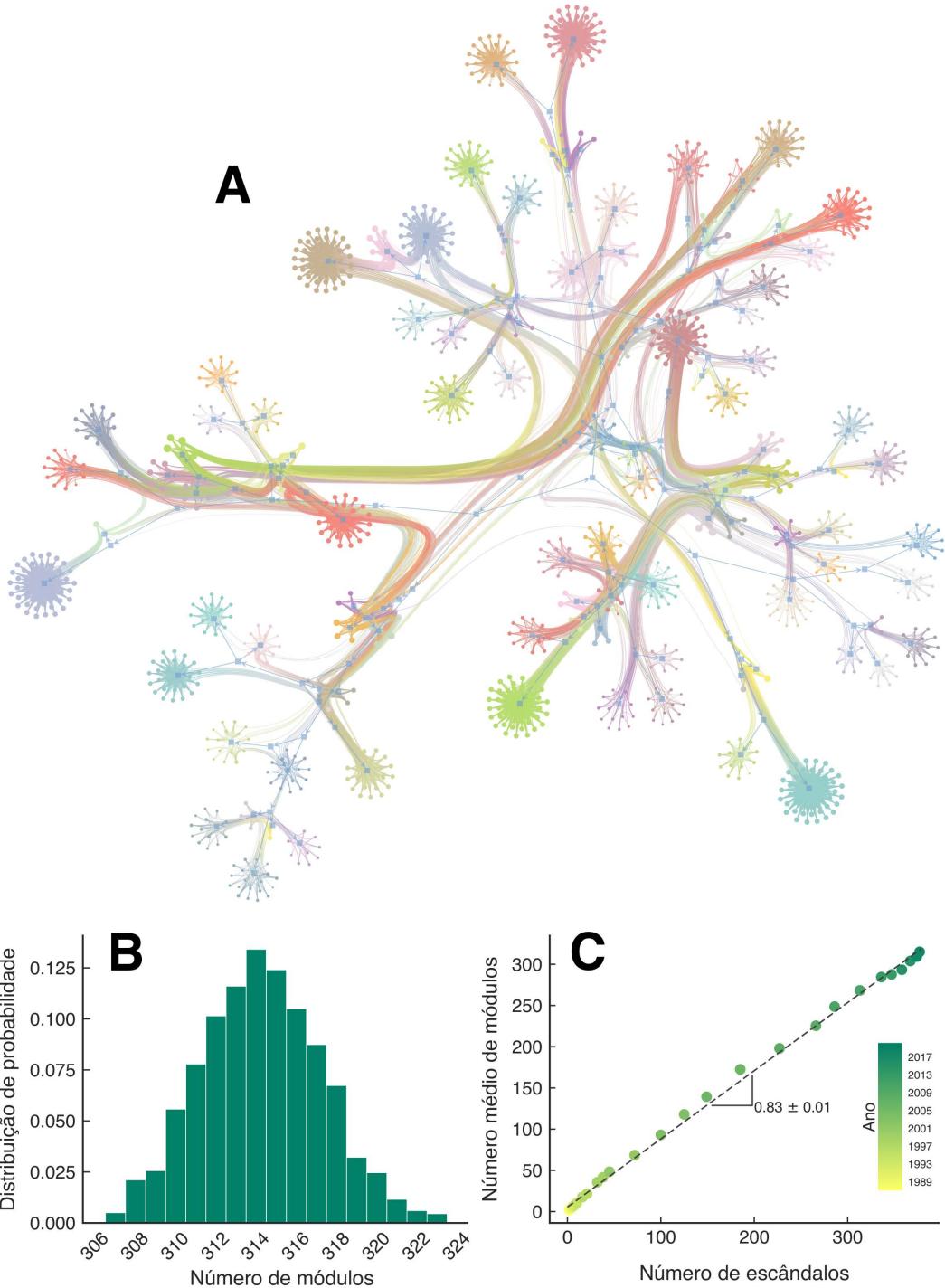


Figura 3.5: (A) Visualização da estrutura modular da maior componente da rede de corrupção política da Espanha. Nessa representação, as diferentes cores e a proximidade de cada nó indicam as estruturas modulares dessa rede obtidas via SBM. (B) Distribuição de probabilidade do número de módulos para a forma final da rede. Essa distribuição foi obtida após 1000 realizações do SBM. (C) Número anual de módulos da rede em função da quantidade de escândalos de corrupção em cada ano. Para cada ano, obtemos uma distribuição de probabilidade similar a do painel (B) e o número de módulos representa a média dessa distribuição. Cada ponto do painel (C) corresponde a um ano diferente e esses pontos são coloridos de acordo com o código de cores posicionado à direita. A linha tracejada representa uma reta com coeficiente angular e erro padrão (indicados na figura) obtidos por meio de uma regressão linear usando o método dos mínimos quadrados [53,54].

existem setas e pequenos quadrados azuis. Essas setas representam uma hierarquia descendente de comunidades: uma espécie de árvore na qual o “topo” representa a maior comunidade e ao “descer” ao longo dessa árvore, encontramos comunidades dentro de comunidades (obtidas pelo modelo SBM aninhado). De forma geral, embora essa figura seja visualmente complexa, ela indica que não somente existe uma estrutura de comunidades na rede de corrupção da Espanha, mas que essa estrutura apresenta uma hierarquia de comunidades.

Uma vez que estamos interessados na quantidade de módulos obtidos para essa rede, é importante salientar que o modelo SBM é intrinsecamente aleatório. Portanto, a Figura 3.5A é apenas uma realização desse modelo, ou seja, o modelo pode retornar um número de comunidades ligeiramente diferente a cada realização. Isso significa que existe mais de uma possibilidade que satisfaça as restrições impostas pelo modelo. Considerando esse aspecto, podemos fazer várias realizações do modelo, obter a distribuição de probabilidade do número de módulos e usar sua média como indicador do número de módulos da rede.

A Figura 3.5B mostra a distribuição de probabilidade do número de módulos obtidos para a rede completa. Essa distribuição foi obtida após 1000 realizações do SBM e indica que a rede de corrupção da Espanha apresenta entre 306 e 324 grupos com valor médio de 314 grupos. Usando essa informação, podemos traçar um paralelo interessante com a rede de corrupção brasileira, a qual contém 27 módulos, 65 casos e, portanto, uma razão de módulos por escândalo de $27/65 \approx 0.42$. Desconsiderando os escândalos que possuem apenas um envolvido, essa razão é quase o dobro para a rede de corrupção espanhola ($187/379 \approx 0.83$). Isso indica que a rede de corrupção espanhola possui um número menor de escândalos dentro de cada comunidade.

Para verificar se o número de módulos possui alguma tendência de evolução, investigamos a relação entre o número de módulos e a quantidade de casos ao longo do tempo, conforme mostra a Figura 3.5A. O número de módulos em um dado ano é uma média da distribuição de probabilidade do número de módulos (obtida de forma similar à Figura 3.5B). Nessa figura, ajustamos um modelo linear e encontramos um bom acordo com os dados.

O coeficiente angular dessa tendência linear é 0.83 e igual à razão obtida para a rede completa. Esse resultado mostra que a razão entre o número de módulos da rede e o número de casos se mantém praticamente constante ao longo de todo o crescimento da rede, sugerindo a existência de um comportamento universal para a evolução de redes de corrupção. Considerando a complexidade envolvida na dinâmica da rede de corrupção espanhola, é bastante surpreendente que sua estrutura de comunidades conserve a razão entre o número de escândalos e a quantidade de módulos.

Outra questão interessante sobre as comunidades dessa rede tem relação com o papel da polarização partidária na formação dessas estruturas. Para analisar esse papel, aplicamos

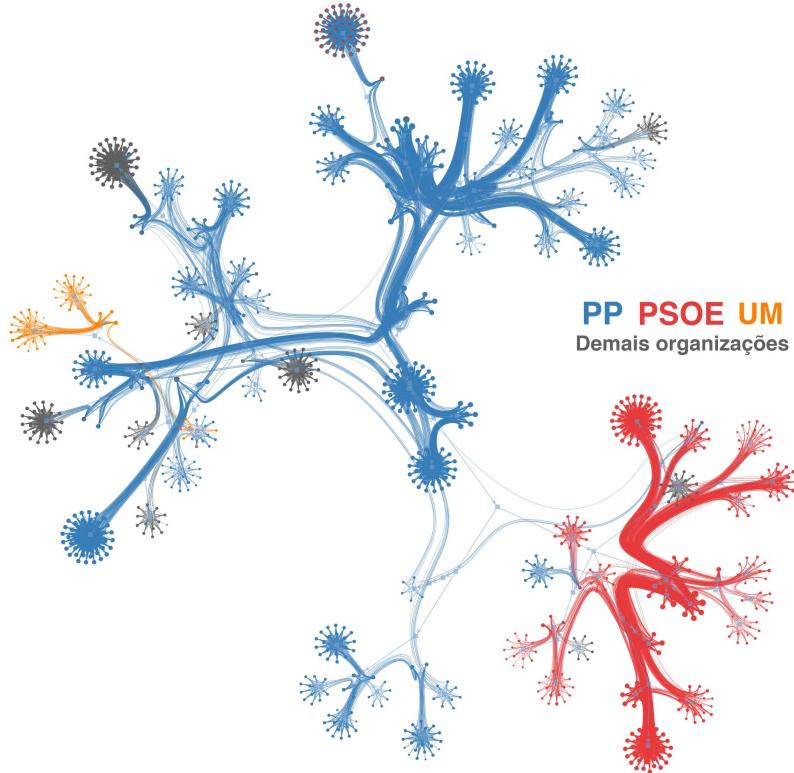


Figura 3.6: Visualização da estrutura modular da componente principal da rede de corrupção espanhola com destaque para as organizações relacionadas aos envolvidos. Os poucos indivíduos que estiveram associados a mais de uma organização possuem vértices divididos igualmente nas cores de cada organização. Podemos observar claramente uma separação das comunidades em dois grandes grupos associados aos dois grandes partidos políticos espanhóis. Dentro desses módulos, envolvidos em atividades criminosas formam estruturas hierárquicas compostas majoritariamente por um único partido (PP ou PSOE).

o modelo SBM utilizado anteriormente à maior componente da rede e colorimos os vértices de acordo com os partidos políticos. A Figura 3.6 apresenta uma dessas realizações. Essa figura mostra claramente uma estrutura de comunidades bipartidária e hierárquica, isto é, as comunidades formadas por envolvidos associados aos grandes partidos espanhóis (PP e PSOE) praticamente não se conectam. Podemos observar dois grandes módulos, cada um associado a um dos dois partidos. Dentro desses módulos, encontramos diversas comunidades formadas por envolvidos em atividades ilícitas associadas ao mesmo partido.

Em linhas gerais, a Figura 3.6 reforça a ideia de que redes criminosas formam comunidades com o objetivo de ocultar suas atividades ilícitas e dificultar o vazamento de informações. Considerando a polarização partidária da rede de corrupção espanhola, podemos verificar que os envolvidos associados aos partidos PP e PSOE estabelecem comunidades hierarquicamente organizadas, o que possivelmente pode ter contribuído para tornar suas atividades mais ocultas, já que são poucos os agentes responsáveis pelas ligações entre as comunidades dos dois partidos.

Capítulo 4

Reincidência criminosa e um modelo para redes de corrupção

Neste capítulo, voltamos nossa atenção ao papel dos envolvidos reincidentes em atividades de corrupção. Primeiramente, lembramos que esses vértices fazem a conexão entre dois ou mais escândalos e, portanto, são os principais responsáveis pela estrutura da rede de corrupção espanhola. Mostramos que envolvidos reincidentes possuem as maiores medidas de centralidade e estão associados às maiores vulnerabilidades da rede com relação a sua fragmentação. Nosso estudo também revelou que o processo tipo coalescência da rede de corrupção espanhola depende de um pequeno número de envolvidos reincidentes. Considerando esses aspectos, a reincidência criminosa se mostra fundamental para coesão e evolução de redes de corrupção.

Motivados por nossos resultados, propomos um modelo computacional para o crescimento de redes de corrupção que se baseia na dinâmica de envolvidos reincidentes. Por uma perspectiva estrutural, os envolvidos reincidentes formam o “esqueleto” da rede e os demais agentes a complementam. Esses outros agentes também são importantes (tal como a categoria de assessores políticos, a qual mostramos atuar de forma mais interna na rede). Entretanto, vértices recorrentes são os principais responsáveis pelo crescimento e estrutura “macroscópica” da rede.

4.1 Dinâmica de reincidência

Iniciamos nossa investigação sobre a reincidência criminosa lembrando que vértices recorrentes da rede são aqueles associados a indivíduos envolvidos em mais de um caso de corrupção. Assim, para cada ano da rede de corrupção espanhola, calculamos a quantidade de vértices recorrentes. A Figura 4.1A mostra o número acumulado de vértices recorrentes ao longo dos 30 anos de nossa base de dados. Analisando os incrementos dessa curva, podemos observar um comportamento semelhante ao da Figura 1.10A, ou seja, o período

de maior variação no número de reincidentes coincide com o período de maior número de envolvidos por ano. Dessa forma, o número de envolvidos reincidentes parece crescer concomitantemente ao total de envolvidos.

Para verificar se isso realmente ocorre e se existe uma relação entre essas duas quantidades, criamos a Figura 1.10B. Essa figura mostra o número acumulado de envolvidos reincidentes (eixo y) em função do número total de envolvidos (eixo x) presentes na rede para cada ano. Nessa figura, cada ponto corresponde a um ano diferente e a linha tracejada representa a reta ajustada por regressão linear usando o método dos mínimos quadrados. Essa figura também mostra os coeficientes angular (0.090) e linear (-11.541) da reta. Encontramos um bom acordo entre dados e modelo linear, indicando que existe uma proporção bastante específica entre a quantidade de vértices recorrentes e o total de envolvidos na rede de corrupção espanhola.

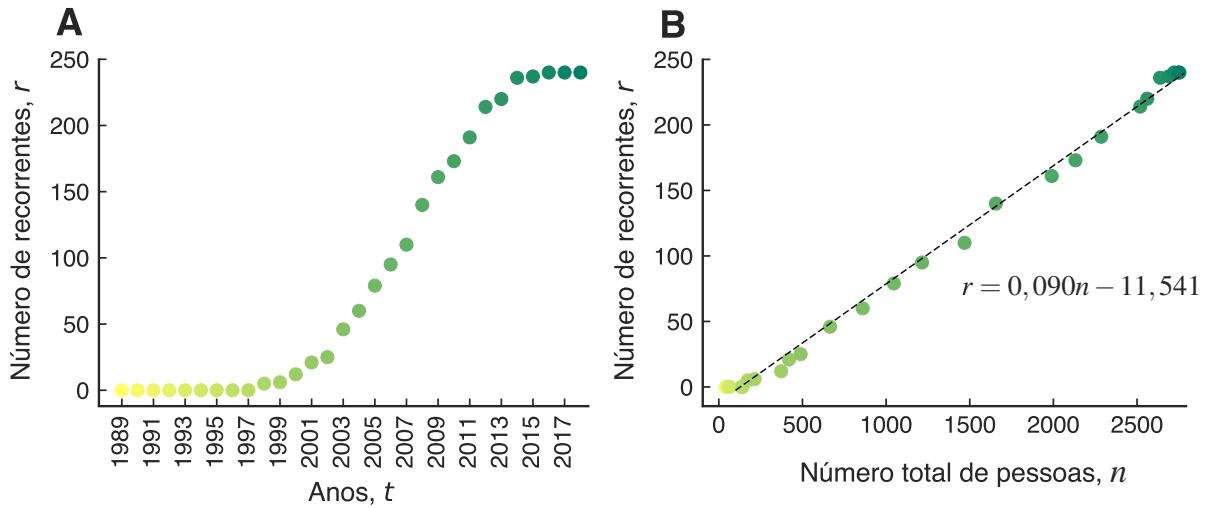


Figura 4.1: (A) Evolução temporal do número acumulado de envolvidos reincidentes presentes na rede de corrupção espanhola. (B) Mesma quantidade do painel anterior em função do número total de envolvidos na rede. Cada ponto dessa figura corresponde a um ano diferente e a linha tracejada representa um ajuste linear aos dados obtido via método dos mínimos quadrados. Esse ajuste apresenta um coeficiente angular igual a 0.090 ± 0.001 e um coeficiente linear igual a -11.541 ± 1.599 .

A Figura 4.1B evidencia que, de fato, existe um padrão no surgimento de vértices recorrentes: o número desses vértices cresce linearmente com o número total de envolvidos. Em média, 9% dos indivíduos envolvidos em novos casos de corrupção são reincidentes na prática criminosa. Além disso, essa fração parece ser constante durante o crescimento da rede. Considerando esse aspecto, a modelagem da rede de corrupção espanhola se mostra ainda mais convidativa. Dado esse padrão, um modelo simplificado pode levar em conta apenas dois aspectos: a proporção de reincidentes por total de vértices e a quantidade típica de envolvidos por escândalos.

Assim, podemos imaginar um modelo que, a cada iteração, adiciona um grafo com-

pleto de um dado tamanho representando um escândalo de corrupção. Grafos completos possuem todos os vértices conectados, da mesma forma que os casos ou escândalos das redes empíricas. O tamanho desses grafos pode ser aproximado pela distribuição exponencial com tamanho característico de ≈ 7 envolvidos por caso, comportamento encontrado para os dados empíricos (Figura 1.10C). Ao passo que o número de vértices aumenta, o número de reincidentes aumenta a uma taxa α de recorrentes por total de vértices. Nesse modelo, para definir os vértices recorrentes, escolhemos aleatoriamente dois vértices não recorrentes para se tornarem um mesmo vértice (ou seja, para desempenhar o papel de reincidente). Portanto, esses vértices conectam duas componentes distintas e, com a adição de novos grafos a cada iteração, a rede evolui e uma estrutura se desenvolve.

O número de recorrentes no modelo pode ser definido pela equação

$$r(n) = \alpha n + \beta, \quad (4.1)$$

na qual r é a quantidade de vértices recorrentes em função do número total de envolvidos n , os quais crescem a cada iteração. Nessa equação, α é a constante representando a taxa de vértices recorrentes e β é outro parâmetro que controla o momento no qual os vértices recorrentes começam a surgir. Esses parâmetros α e β estão diretamente relacionados aos coeficientes angular e linear da reta mostrada na Figura 4.1B. O ajuste linear da Figura 4.1B indica que $\alpha \approx 0.09$ e $\beta \approx -11.541$. Assim, para que o número de reincidentes seja maior do que zero ($r(n) > 0$) é preciso que a rede de corrupção espanhola possua cerca de $11.541/0.090 \approx 129$ vértices.

De forma geral, o aspecto mais importante desse modelo é a taxa de envolvidos reincidentes α . Esse parâmetro controla a proporção de envolvidos reincidentes na rede e, portanto, seu valor deve alterar substancialmente a estrutura da rede gerada. Embora tenhamos encontrado um valor típico ($\alpha \approx 0.090$) para essa taxa durante toda a evolução da rede de corrupção espanhola, o estudo do modelo com diferentes valores para α pode fornecer informações importantes. Esse estudo pode revelar o papel desempenhado por esse tipo de vértice na estrutura das redes formadas pelo modelo. Na próxima seção, apresentamos uma análise na qual fixamos os outros parâmetros do modelo e variamos apenas α .

Contudo, antes de prosseguirmos, um detalhe merece destaque. Na definição atual do nosso modelo, apenas vértices não recorrentes podem se tornar recorrentes. No entanto, podemos imaginar que existem envolvidos reincidentes em mais de dois casos de corrupção. De fato, embora aproximadamente apenas 2.5% de todos os indivíduos no estágio final da rede satisfaçam esse critério, podemos considerar esse valor como um parâmetro ($p_a = 0.025$) adicional do modelo. Incluímos esse comportamento em nosso modelo de modo que, ao selecionar um vértice recorrente, existe uma probabilidade p_a dele já ter sido recorrente anteriormente. Em nossas análises, observamos que esse parâmetro afeta pouco a estrutura

da rede, principalmente, se $p_a < 0.1$.

De modo mais detalhado, o algoritmo abaixo mostra os passos do modelo:

- Inicializar uma rede vazia ($n = 0$);

Repetir um dado número de vezes:

- Gerar um grafo totalmente conectado de tamanho sorteado de uma distribuição exponencial;
- Adicionar o grafo à rede;
- Definir o número de novos agentes recorrentes usando o total de agentes recorrentes $r(n) = \alpha n + \beta$;

Se o número de novos agentes recorrentes for maior zero:

Repetir para cada novo agente recorrente:

- Com probabilidade $1 - p_a$, escolher aleatoriamente dois vértices não recorrentes para se tornarem um mesmo vértice recorrente;
- Com probabilidade p_a , escolher aleatoriamente um vértice recorrente e um vértice não recorrente para se tornarem um mesmo vértice recorrente;

Consideramos que existem muitas outras análises interessantes sobre o comportamento detalhado desse modelo. No entanto, em nosso trabalho, estudamos apenas o comportamento desse modelo em função da taxa de reincidência criminosa α . Conforme veremos, essa escolha é suficiente para entender e modelar vários aspectos da rede real.

4.2 Comportamento do modelo computacional

Considerando os aspectos discutidos anteriormente, verificamos que existe a possibilidade de modelar o crescimento de redes de corrupção utilizando ingredientes simples, tais como a taxa de reincidência α criminosa e a quantidade típica de envolvidos por escândalos de corrupção ($1/\lambda$). Nossos achados indicam que essa taxa α permanece constante ao longo de toda a evolução da rede. Além disso, nossos resultados mostram que o número de envolvidos por caso de corrupção segue uma distribuição exponencial com número típico de indivíduos $1/\lambda \approx 7.33$ (Figura 1.10C).

Assim, o comportamento do modelo perante a variação de α (equação 4.1) pode produzir comportamentos interessantes. Uma vez que α varia entre 0 e 1, é pertinente saber o comportamento das redes geradas nesses casos limites. De forma mais específica, nossa investigação sobre o comportamento do modelo será realizada estudando o tamanho da maior componente das redes geradas para cada valor de α . A cada iteração do modelo, um grafo completo com um dado número de vértices (ditado pela distribuição exponencial com $1/\lambda = 7.33$) é adicionado à rede. Assim, o tamanho n da rede aumenta e a quantidade de envolvidos reincidentes $r(n)$ cresce segundo a equação 4.1.

A Figura 4.2 mostra a média dos tamanhos da maiores componentes das rede obtidas via nosso modelo em função de α . Para cada valor diferente de α , crescemos a rede por meio de 400 iterações do modelo¹ e extraímos a fração da maior componente da rede gerada pela última iteração. Por fim, repetimos esse processo 1000 vezes e calculamos uma média do tamanho da maior componente para cada um dos valores de α . A banda colorida (em verde marinho claro) representa os valores máximo e mínimo do tamanho da maior componente nas 1000 realizações correspondente a cada valor de α . Aqui, usamos os valores máximos e mínimos porque os erros padrões das médias são muito pequenos e não ficam visíveis na figura. Isso significa que as médias variaram pouco para cada valor de α .

O gráfico inserido bem no centro da Figura 4.2 corresponde à derivada da curva citada anteriormente. Construímos essa derivada para destacar a variação brusca ocorrida nessa curva em torno de $\alpha \approx 0.073$. Podemos notar que o gráfico apresenta um ponto máximo nesse valor e, como veremos adiante, esse comportamento corresponde a uma espécie de transição de fase nas redes formadas pelo modelo.

As visualizações das redes inseridas na Figura 4.2 representam formas típicas das redes geradas para três valores diferentes de α . Observamos que se o valor de α é muito pequeno, a rede formada é esparsa e composta por um grande número de componentes isoladas. No outro extremo, valores de α próximos de 1 produzem redes muito conectadas e que apresentam uma estrutura em forma de “corrente”. Entretanto, esse modelo apresenta uma espécie de transição de fase entre esses dois extremos ao redor de $\alpha = 0.073$ (ponto crítico do modelo). Curiosamente, a rede real de casos de corrupção espanhola apresenta uma taxa de reincidência criminosa ($\alpha = 0.090$) bem próxima desse ponto crítico do modelo. Na Figura 4.2, a linha contínua em preto indica a taxa de reincidência que coincide com o ponto crítico do modelo (α_m) e a linha tracejada em vermelho indica o valor da taxa de reincidência (α_e) obtida empiricamente pela reta da Figura 4.1B.

¹Verificamos que o número de vértices das redes simuladas é aproximadamente igual ao da rede real após 400 iterações. Se a distribuição de agentes por caso fosse exatamente uma exponencial com número característico de pessoas igual a 7.33, seriam necessários, em média, $2753/7.33 \approx 375$ passos para gerar redes com 2753 pessoas (tamanho da rede real). Porém, como a distribuição exponencial subestima a distribuição empírica do número de agentes por caso (Figura 1.10C), são necessários alguns passos adicionais para aproximar o tamanho da rede real.

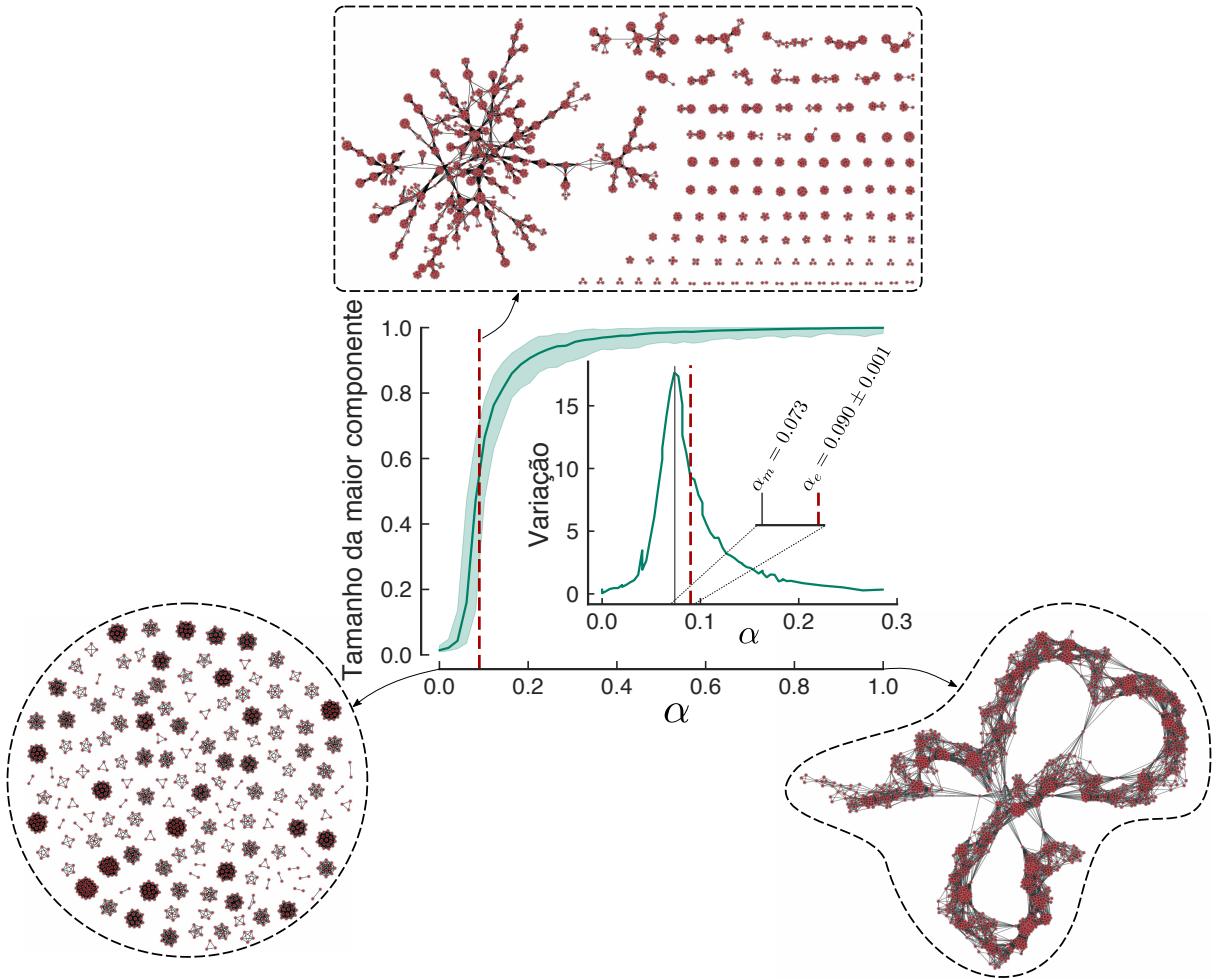


Figura 4.2: Tamanho médio das maiores componentes das redes obtidas via nosso modelo para redes de corrupção em função da taxa de reincidência α . A curva inserida internamente nesse gráfico representa a derivada da quantidade anterior. Nessa curva, podemos observar um valor máximo, correspondendo a uma espécie de transição de fase do modelo. Em ambas as curvas, as linhas verticais tracejadas representam o mesmo ponto ($\alpha = 0.073$) correspondente a essa transição. Curiosamente, o valor empírico obtido para a taxa de reincidência da rede de corrupção espanhola ($\alpha = 0.090$) é muito próximo ao valor $\alpha = 0.073$. Para realçar isso, uma inserção no gráfico mostra a diferença entre os valores obtidos empiricamente (α_e) e por meio do modelo (α_m). As três redes mostradas na figura correspondem às formas típicas das redes geradas para $\alpha = 0$, $\alpha = 0.073$ e $\alpha = 1$.

Esses três valores típicos para a taxa de reincidência α podem ser entendidos da seguinte maneira:

- Quando $\alpha = 0$, o número de vértices recorrentes é igual β e não aumenta com o crescimento da rede. Assim, se β for pequeno ou negativo (como em nosso caso), as redes simuladas crescem via adição grafos completos isolados e a estrutura final é composta de um grande número de componentes isoladas com tamanho típico igual ao tamanho típico dos casos de corrupção. Um exemplo desse caso é mostrado à esquerda da Figura 4.2.

- Quando $\alpha = 1$, quase todos os novos vértices adicionados à rede em cada iteração se tornam recorrentes. Nesse caso, a rede cresce produzindo uma estrutura em forma de “corrente”, conforme mostrado à direita na Figura 4.2.
- Quando $\alpha \approx 0.073$ (ponto crítico do modelo), observamos um comportamento similar ao da rede empírica, com a presença de uma grande componente conectada e outras componentes isoladas menores. Como mencionamos, esse valor crítico de α é muito próximo do obtido empiricamente para a rede de corrupção espanhola (0.090). Portanto, a rede de corrupção espanhola parece operar muito próximo desse ponto crítico, abaixo do qual ocorre a fragmentação da rede e acima do qual a rede se torna muito conectada. Surpreendentemente, esse valor permanece praticamente constante em toda a evolução da rede de corrupção espanhola.

Nesse trabalho, investigamos o comportamento do modelo apenas com relação ao parâmetro α . Porém, acreditamos que outros estudos que levem em conta os demais parâmetros e considerem algumas outras variações (como a distribuição de tamanho dos casos) podem ser bastante interessantes e são uma perspectiva para trabalhos futuros. Nessa dissertação, visto que o caso $\alpha \approx 0.073$ gera redes semelhantes às redes empíricas, nos concentramos em comparar o comportamento da rede de corrupção espanhola com as redes geradas pelo modelo próximas a esse ponto crítico. A próxima seção apresenta essa análise.

4.3 Comparação com resultados empíricos

Nossos achados sobre redes de corrupção permitiram propor um modelo baseado na taxa de envolvidos reincidentes e no tamanho típico de pessoas por casos de corrupção. Para valores de α próximos a 0.073, nosso modelo gera redes de aparência similar às redes reais de corrupção. Visto que essa taxa é bem próxima à encontrada para a rede de corrupção espanhola, nosso modelo pode ser um bom candidato para descrever diferentes aspectos dessas redes de corrupção. Portanto, podemos verificar se o comportamento das redes geradas pelo modelo se assemelha ao comportamento da rede de corrupção espanhola.

Para comparar o crescimento entre as redes geradas pelo modelo e a rede de corrupção espanhola, vamos separar as realizações do modelo em tempos t . Esses tempos possuem uma correspondência aos anos da rede real, ou seja, podemos escrever $t \in \{1989, \dots, 2018\} \rightarrow \{0, \dots, 29\}$. Por exemplo, visto que em 2018 surgiram 2 escândalos na rede de corrupção espanhola, no tempo $t = 29$ serão feitas 2 iterações do modelo e cada uma irá adicionar um conjunto de vértices conectados (os “casos”).

O crescimento das redes via nosso modelo será realizado com os seguintes critérios. O número de casos que entram na rede no tempo t segue a série temporal da Figura 1.10B

e o tamanho de cada caso é obtido a partir da distribuição exponencial da Figura 1.10C. O número de vértices recorrentes cresce de acordo com a equação 4.1 com parâmetros ($\alpha = 0.090$ e $\beta = -11.541$) fixos. Além disso, usamos $p_a = 0.025$, mesmo valor encontrado empiricamente.

Vamos considerar duas propriedades das redes geradas pelo modelo: distribuição de grau e evolução do tamanho da maior componente. Para deixar mais clara a comparação, os painéis (A) e (B) da Figura 4.3 destacam essas duas propriedades da rede de corrupção espanhola. A Figura 4.3A representa uma versão modificada da Figura 4.3B, na qual o grau não é reescalado. Essa modificação será útil para verificar se os valores de grau das distribuições (do modelo e empírica) são parecidos.

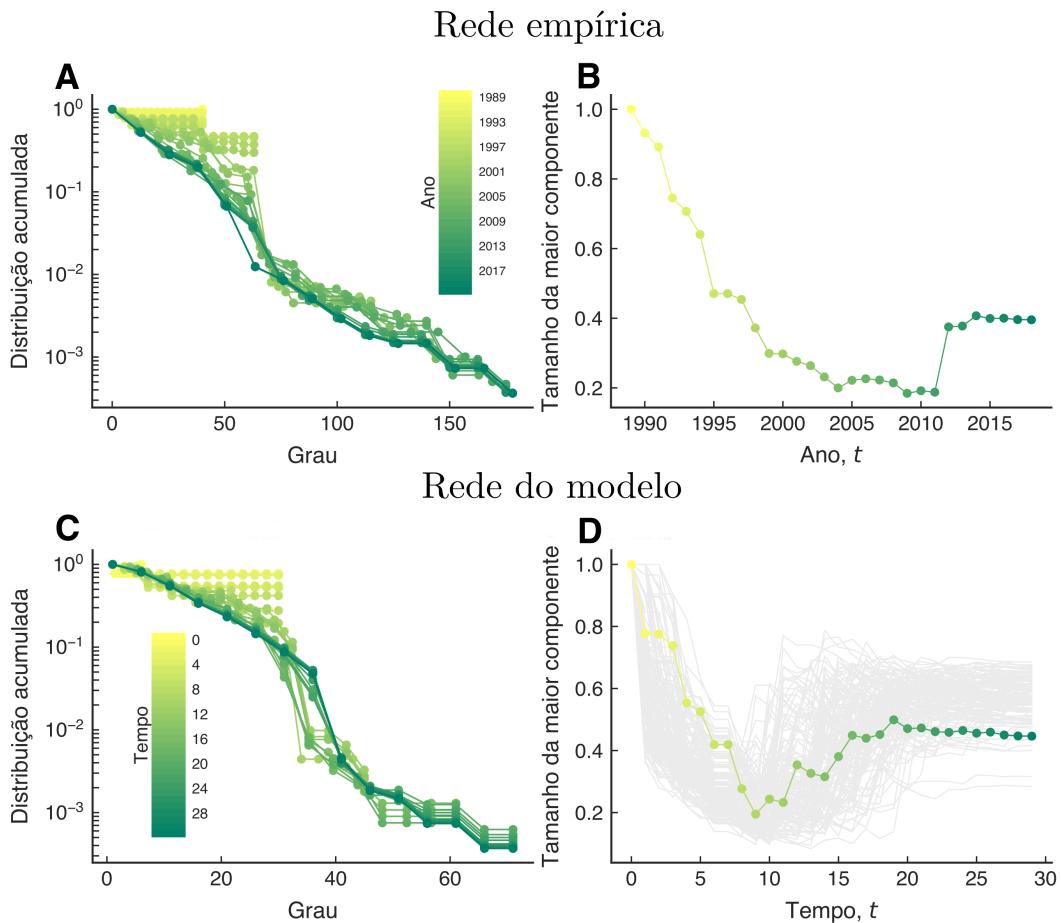


Figura 4.3: Comparação entre as redes geradas pelo modelo e a rede de corrupção espanhola. (A) Evolução temporal da distribuição acumulada de grau e (B) série temporal do tamanho (normalizado) da maior componente da rede de corrupção espanhola. (C) Evolução temporal da distribuição acumulada de grau para uma rede gerada por nosso modelo entre $t = 0$ e $t = 29$. (D) Tamanho da maior componente de 100 realizações das redes geradas por nosso modelo (em cinza claro). A curva destacada representa o comportamento médio sobre todas as realizações do modelo.

Construímos também a Figura 4.3B, a qual mostra o tamanho normalizado da maior componente da rede de corrupção espanhola ao longo de todos os anos de nossa base de

dados. Nesse gráfico, podemos notar uma variação brusca no tamanho da maior componente correspondente ao processo tipo coalescência ocorrido entre 2011 e 2012.

A Figura 4.3C mostra a evolução da distribuição de grau de uma rede típica gerada pelo nosso modelo. A cor de cada curva refere-se a cada tempo t e corresponde ao código de cores mostrado na figura. Visualmente, a evolução da distribuição de grau se assemelha bastante à evolução da distribuição de grau da rede de corrupção espanhola. No entanto, notamos uma diferença nos valores máximos de grau das distribuições para o caso empírico e teórico. A Figura 4.3A possui distribuições com alguns valores de grau maiores quando comparados aos valores de grau presentes na distribuição da Figura 4.3C. Isso ocorre porque a distribuição exponencial do número de pessoas por caso, a qual consideramos para construir o modelo, não se ajusta muito bem aos dados para escândalos muito grandes (Figura 1.10C).

Entretanto, a evolução temporal e a forma das distribuições das redes geradas por nosso modelo possuem bastante semelhança visual com o caso empírico. Além disso, embora essa seja apenas uma realização da evolução de uma rede gerada pelo modelo, outras realizações produzem distribuições de grau com o mesmo padrão visual.

Em relação ao outro comportamento das redes geradas pelo nosso modelo, a Figura 4.3D mostra o tamanho normalizado (isto é, dividido pelo total de vértices da rede) da maior componente para diversas redes simuladas. Nessa figura, as curvas em cinza claro representam 100 realizações de evoluções das redes geradas pelo modelo e a curva em destaque representa o valor médio sobre todas as realizações.

Podemos observar uma grande semelhança entre as curvas das Figuras 4.3B e 4.3D. Embora as curvas não decaiam durante o mesmo intervalo de tempo, ambas decrescem até passar por um mínimo e depois se estabilizam em torno do valor 0.4. Podemos imaginar que essas curvas apresentam um comportamento similar porque a quantidade de casos por unidade de tempo que é adicionada nas redes empírica e teóricas é a mesma (segundo a Figura 1.10B). No entanto, o tamanho desses casos não é exatamente igual; apenas a distribuição de probabilidade da quantidade de pessoas por escândalo é a mesma.

Notamos ainda que as curvas simuladas da Figura 4.3D se estabilizam mais rapidamente do que a curva empírica da Figura 4.3B. Essa diferença pode estar associada à polarização partidária da rede empírica. No caso da curva empírica, a estabilização ocorre apenas após o processo tipo coalescência responsável pela junção das componentes associadas aos dois principais partidos da Espanha. Dessa forma, a polarização partidária presente na rede de corrupção espanhola pode ter sido responsável por uma espécie de atraso na estabilização do tamanho de sua maior componente.

A Figura 4.3D também revela possíveis processos de coalescências nas redes geradas pelo modelo (indicadas por algumas variações abruptas no tamanho da maior componente). Embora não sejam tão bruscas quanto a variação observada entre 2011 e 2012 da rede real (Figura 4.3B), podemos verificar se esses processos realmente ocorrem e se

apresentam uma forma similar ao observado para a rede de corrupção espanhola.

Conforme ilustra a Figura 4.4, encontramos processos tipo coalescência nas redes geradas pelo nosso modelo. A Figura 4.4A apresenta a evolução temporal (de $t = 0$ a $t = 29$) do tamanho das duas maiores componentes para 100 realizações das redes geradas pelo modelo. As várias redes simuladas apresentam variações bruscas no tamanho das maiores componentes. Visto que vértices nunca são removidos da rede, esse gráfico indica processos tipo coalescência nos quais as duas maiores componentes se juntam. Em especial, as cores roxo e verde mais escuras da Figura 4.4A destacam uma realização do modelo que passou por um processo tipo coalescência entre $t = 21$ e $t = 22$.

Para investigar melhor esse comportamento, a Figura 4.3B mostra a variação anual das mesmas quantidades da Figura 4.4A. Aqui as cores roxa e verde mais escuras correspondem à evolução da mesma rede destacada na Figura 4.3A. De fato, observamos um valor máximo para a maior componente e, concomitantemente, um valor mínimo para a segunda maior componente. Esse gráfico identifica precisamente uma transição abrupta ocorrida na rede crescida pelo modelo, isto é, a maior componente aumenta bruscamente de tamanho porque um novo conjunto de vértices faz as duas maiores componentes em $t = 21$ se juntarem em $t = 22$.

Podemos fazer uma terceira análise mais qualitativa observando as componentes graficamente em uma realização do modelo, conforme mostra a Figura 4.4C. Para isso, identificamos as maiores componentes da rede para o tempo $t = 21$ e descobrimos os conjuntos de vértices que as conectam em $t = 22$. Esse conjunto de vértices está representado em vermelho na figura. De fato, é possível confirmar que as duas maiores componentes da rede em $t = 21$ se conectam para formar uma componente gigante em $t = 22$. Esse processo é similar ao comportamento tipo coalescência observado na rede empírica, no qual poucos vértices recorrentes são responsáveis pela conexão entre duas ou mais componentes da rede.

Nossos resultados mostram que é possível modelar o crescimento de redes de corrupção utilizando ingredientes simples, dentre os quais, a taxa de reincidência criminosa (quando um indivíduo já implicado em um escândalo pratica outro delito) mostra-se fundamental para a existência de coalescência e para a maior coesão de redes de corrupção.

Além de fornecer uma regra para o crescimento de redes de corrupção, esse modelo indica que existe uma taxa específica de reincidência criminosa responsável pela formação de redes visualmente similares à rede de corrupção espanhola. Mostramos que o modelo operando com valores de α próximos dessa taxa crítica não apenas produz redes visualmente similares, mas também é capaz de reproduzir alguns comportamentos da rede de corrupção espanhola. Entre eles, a evolução da distribuição de grau da rede e o crescimento de sua maior componente. Por fim, notamos que o modelo reproduz essencialmente o mesmo comportamento do processo tipo coalescência observado nas redes de corrupção brasileira [18] e espanhola.

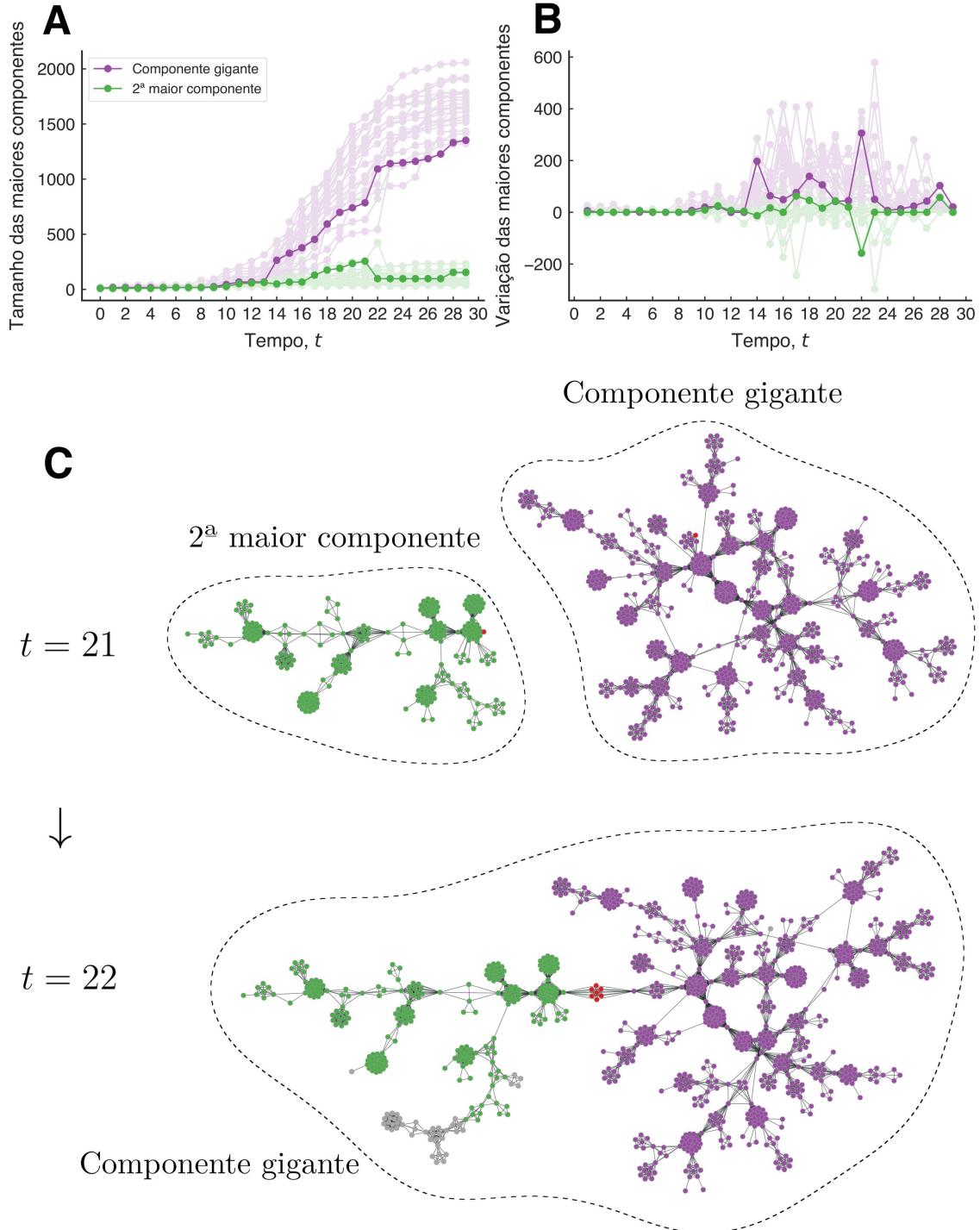


Figura 4.4: Processo tipo coalescência observado nas redes geradas pelo nosso modelo. (A) Evolução temporal do tamanho das duas maiores componentes para várias realizações do modelo. (B) Variação anual do tamanho das componentes do painel anterior. Em ambos os painéis, as cores mais claras representam as diferentes realizações do modelo. Observamos variações bruscas nos tamanhos das duas maiores componentes de forma bastante similar ao que encontramos para a rede de corrupção espanhola. Destacamos uma dessas ocorrências em ambos os painéis anteriores com cores mais escuras. (C) Visualização do processo de coalescência entre as duas maiores componentes de uma rede simulada. De $t = 21$ a $t = 22$ alguns poucos agentes (em vermelho) presentes na rede em $t = 21$ se tornam recorrentes no tempo $t = 22$. Isso faz com que ocorra a junção entre as duas maiores componentes da rede do tempo $t = 21$ no tempo $t = 22$.

Conclusões e perspectivas

Nesse trabalho, desenvolvemos um estudo sobre redes de corrupção. Para realizar essa investigação, utilizamos uma base de dados com informações sobre escândalos de corrupção ocorridos na Espanha entre 1989 e 2018. Nossas análises foram baseadas em ciência de redes e comparamos nossos achados com resultados da literatura sobre uma rede de corrupção brasileira [18].

Apresentamos informações demográficas sobre esses dados e destacamos o papel do bipartidarismo político espanhol nos escândalos de corrupção e seus envolvidos. Encontramos um número típico (parecido com o caso brasileiro) para a quantidade de envolvidos por escândalo de corrupção, reforçando a hipótese de que a maioria dos envolvidos tendem a se envolver em grupos pequenos para evitar a exposição das atividades ilícitas.

Ao construir a rede de corrupção espanhola, notamos uma separação clara entre os envolvidos associados aos dois maiores partidos políticos espanhóis (PP e PSOE). Estudamos algumas medidas estruturais dessa rede e encontramos similaridades com a rede de corrupção brasileira. Mostramos que a rede de corrupção espanhola possui medidas de assortatividade e caminho médio maiores, indicando possíveis consequências da polarização política presente nessa rede. Aplicamos medidas de centralidade aos vértices da rede levando em consideração seus gêneros e categorias. Futuros trabalhos podem considerar a rede de corrupção espanhola como uma rede multicamadas, na qual cada camada representa um cargo diferente. Essa abordagem pode oferecer mais informações sobre as interações entre os diferentes cargos.

Em outra análise, examinamos a vulnerabilidade da rede de corrupção espanhola por meio de seu desmantelamento com ataques direcionados a vértices de categorias específicas. Mais análises nesses dados podem mostrar se ataques a vértices que conectam dois ou mais módulos são mais eficazes.

Em seguida, levamos em consideração o aspecto temporal da rede de corrupção espanhola. Estudamos a evolução de sua distribuição de grau e encontramos que a distribuição exponencial aproxima os dados razoavelmente bem. Após isso, estudamos a evolução temporal das duas maiores componentes da rede de corrupção espanhola. Mostramos que essa rede passou por uma transição abrupta por consequência da junção de suas três maiores componentes associadas a três partidos diferentes. Esse fenômeno tipo coalescência ocorreu de forma parecida na rede de corrupção brasileira, no sentido de que alguns poucos

envolvidos reincidentes foram responsáveis por conectar diferentes componentes.

Encontramos também os módulos que formam a rede de corrupção espanhola. Analisamos sua estrutura de comunidades e verificamos que essa rede possui uma estrutura hierárquica de comunidades associada aos dois maiores partidos espanhóis. Além disso, encontramos um padrão linear na evolução temporal do número de casos por módulo. Assim, mesmo possuindo escândalos com vários desdobramentos, a rede de corrupção espanhola apresenta um número menor de escândalos por módulo quando comparada a rede de corrupção brasileira.

Por fim, propomos um modelo computacional simples para descrever o crescimento de redes de corrupção. Nosso modelo é baseado em dois ingredientes principais: reincidência criminosa e número típico de envolvidos por escândalo. Verificamos que esse modelo apresenta uma espécie de transição de fase para um valor específico da taxa de reincidência. Esse valor crítico é muito próximo da taxa de reincidência da rede real. Ao estudar as redes simuladas ao redor desse ponto crítico, verificamos que o modelo é capaz de reproduzir diversos resultados empíricos da rede de corrupção espanhola.

Acreditamos que existem muitas outras possibilidades para explorar nosso modelo em trabalhos futuros. Dentre as mais diretas, destacamos uma investigação mais completa de todos seus parâmetros e estudos sobre como esses parâmetros afetam, por exemplo, a estrutura modular das redes simuladas. Além disso, podemos verificar se esse mesmo modelo é capaz de reproduzir características de outras redes reais de corrupção (como a brasileira [18]) ou, de modo mais geral, descrever outras redes criminosas.

De um ponto de vista mais aplicado, nossas análises revelaram que a reincidência criminosa desempenha um papel crucial para a estrutura e dinâmica de redes de corrupção. De acordo com nosso modelo, para reduzir a coordenação entre agentes em redes de corrupção, parece ser fundamental evitar a reincidência criminosa de indivíduos. Parte dessa redução poderia ser alcançada por meio da aplicação mais efetiva das medidas cautelares e de privação de liberdade previstas em lei. A generalização desses resultados pode ser útil para guiar políticas públicas e medidas de combate à corrupção.

Apêndice A: Conceitos de ciência de redes

A.1 Representação de redes

Uma rede (formalmente conhecida como grafo) é caracterizada por um par ordenado $G = (V, E)$, no qual V representa o conjunto dos vértices (ou nós) e E define o conjunto composto pelas relações entre esses vértices. Designamos a letra N para representar o número total de vértices e a letra L para caracterizar o número de ligações (ou arestas) entre esses vértices. Embora essa seja a definição mais simples de uma rede, seus componentes fundamentais (vértices e ligações) constituem a base mais importante de toda nossa abordagem.

De forma geral, um grafo é completamente definido por uma matriz de adjacência \mathbf{A} . Um grafo simples possui uma matriz de adjacência \mathbf{A} com os elementos a_{ij} iguais a 1 caso os vértices i e j estejam conectados e 0 caso contrário. A Figura A1 mostra uma matriz de adjacência com sua respectiva representação em forma de vértices e ligações.

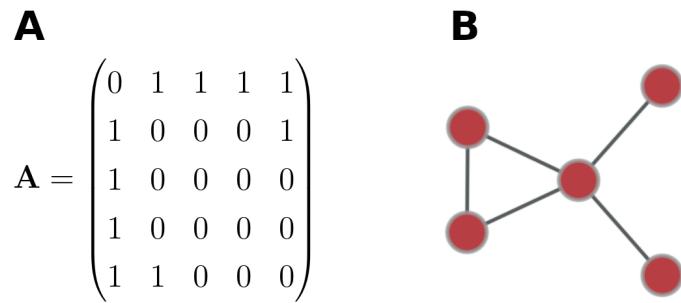


Figura A1: (A) Matriz de adjacência na qual cada elemento a_{ij} indica se há conexão entre os vértices i e j da matriz. (B) Visualização do grafo formado pela matriz de adjacência do painel (A).

A.2 Medidas estruturais

A.2.1 Coeficiente de agrupamento

O coeficiente de agrupamento [22] de um dado vértice i representa o quanto conectado seus vizinhos são. Se L_i representa o número de ligações entre os k_i vizinhos de i , então

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (\text{A.1})$$

é o coeficiente de agrupamento do vértice i . Além disso, podemos tomar a média dos coeficientes de agrupamento para uma rede de tamanho N . Definida como

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i, \quad (\text{A.2})$$

essa medida representa o coeficiente de agrupamento global da rede.

A.2.2 Assortatividade

O coeficiente de assortatividade r [55] de uma rede calcula a tendência de vértices com grau similar se conectarem. Essa medida é definida como o coeficiente de correlação de Pearson do grau entre pares de vértices conectados.

Obtemos $r = 0$, quando não há tendência de vértices de grau similar se conectarem. Por outro lado, $r = 1$ quando todos os nós de grau similar se conectam. No caso completamente oposto, em que todos os vértices se conectam com vértices de grau diferente, temos $-1 \leq r < 0$.

A.2.3 Comprimento médio do caminho

Denotado por $\langle d \rangle$, o comprimento médio do caminho [22] de uma rede é a média da soma dos caminhos mais curtos entre todos os vértices da rede. Para uma rede de tamanho N ,

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i \neq j}^N d_{i,j}, \quad (\text{A.3})$$

sendo $d_{i,j}$ a distância entre os vértices i e j .

A.3 Medidas de centralidade

A.3.1 Grau

A centralidade de grau [22] oferece uma das grandezas mais básicas de redes. No caso mais simples, o grau de um determinado vértice representa seu número de ligações. Podemos obter o grau k_i de um vértice i pela matriz de adjacência da rede, isto é,

$$k_i = \sum_{j=1}^N A_{ij}, \quad (\text{A.4})$$

sendo N o número de vértices da rede.

A.3.2 Intermediação

A centralidade de intermediação [56] mede o número de vezes no qual um vértice se encontra no caminho de outros vértices. Essa medida identifica os vértices que atuam como elos na rede, fornecendo um valor entre 0 e 1. Mais especificamente, a centralidade de intermediação de um vértice i é definida como

$$C_B(i) = \frac{2}{(N-1)(N-2)} \sum_{s,t \in V} \frac{\sigma_{st}(i)}{\sigma_{st}}, \quad (\text{A.5})$$

com $s \neq i \neq t \neq s \in V$. Aqui, designamos o número de caminhos entre dois vértices quaisquer por σ_{st} e aos que passam pelo vértice i por $\sigma_{st}(i)$. Sendo N o número de vértices da rede, $\frac{(N-1)(N-2)}{2}$ representa o número de possíveis caminhos que passam pelo vértice i em uma rede não direcionada. Esse último fator normaliza a medida de centralidade de intermediação para o vértice i .

A.3.3 Proximidade

A centralidade de proximidade [56] retorna um valor para cada vértice baseado em sua distância em relação a todos os outros vértices da rede. Para um dado vértice i , essa medida é definida como

$$c_i = \frac{1}{\sum_j d_{ij}}, \quad (\text{A.6})$$

sendo d_{ij} a menor distância entre os vértices i e j .

A.4 Distribuição de grau

Na sua forma mais simples, a distribuição de grau de uma rede representa a fração de vértices que possuem um número de conexões igual a k . Para uma rede com N vértices, a probabilidade de encontrar um vértice com grau k é dada por $P(k) = \frac{N_k}{N}$, sendo N_k o número de vértices com grau k . A Figura A2 mostra uma distribuição de grau típica.

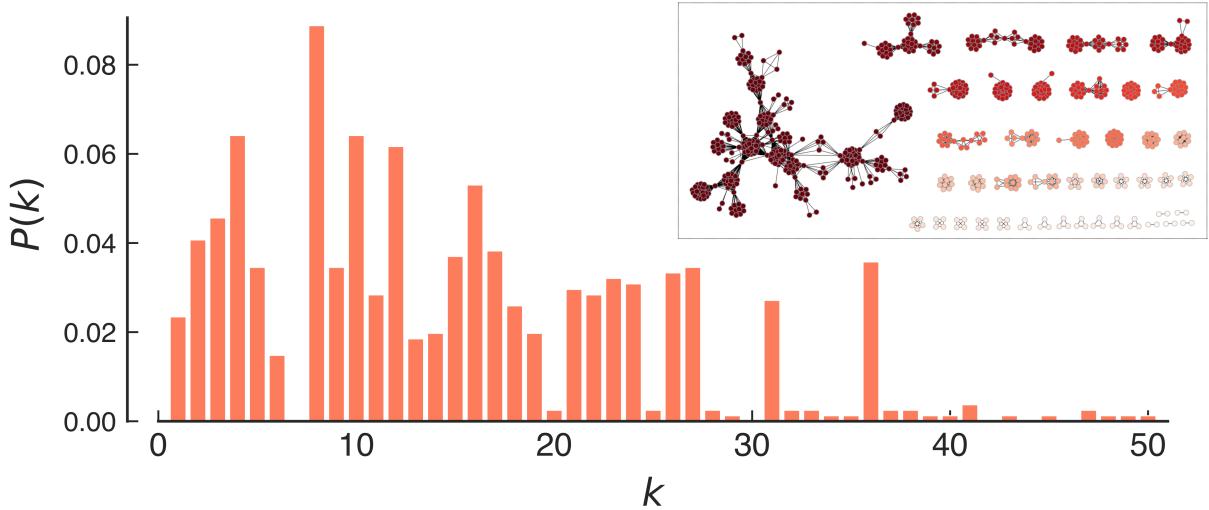


Figura A2: Exemplo de distribuição de grau para a rede com N vértices mostrada na inserção à direita.

Outra maneira comum de representar a distribuição de grau é por meio da distribuição acumulada complementar². Nessa representação, consideramos a fração de vértices que possuem grau maior do que um valor k . A Figura A3 mostra a distribuição acumulada complementar da mesma rede ilustrativa usada na Figura A2.

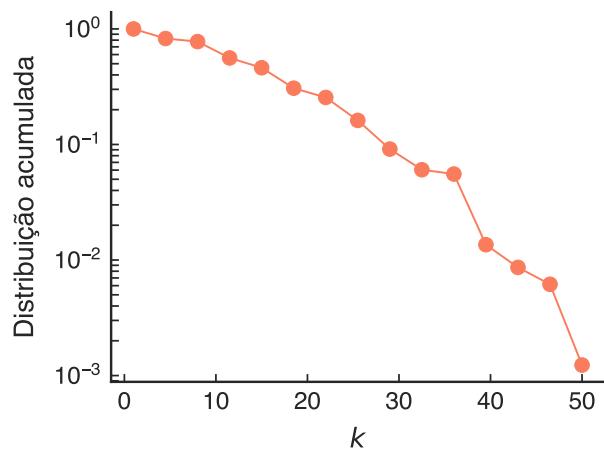


Figura A3: Distribuição de grau acumulada (complementar) da rede da Figura A2.

²O Apêndice B.3 descreve em maior detalhe as distribuições acumulada e acumulada complementar.

A.4 Modelo de Erdös-Rényi

Um dos principais modelos para criar redes aleatórias é conhecido como modelo de Erdös-Rényi [57]. Uma das definições desse modelo considera uma função $G(n, m)$ para representar um grafo escolhido uniformemente ao acaso a partir da coleção de todos os grafos que possuem n vértices e m ligações. Assim, um grafo $G(n, m)$ apresenta $\binom{\binom{n}{2}}{m}$ possibilidades e cada possibilidade ocorre com probabilidade $1/\binom{\binom{n}{2}}{m}$.

A.5 Modelo bayesiano de blocos estocásticos

Embora não exista uma descrição absoluta para definir o que é uma comunidade (ou módulo), de forma simplificada, podemos considerar como comunidades conjuntos de vértices que estão mais conectados entre si do que com outros vértices. Para resolver o problema de encontrar esses conjuntos de vértices, inúmeros métodos já foram propostos [58]. Um desses métodos é conhecido como modelo de blocos estocásticos (do inglês, *Stochastic Block Model* ou SBM) [59]. Esse modelo é gerativo [60] e permite a inferência de informações a partir de dados de redes reais.

De forma geral, o modelo de blocos estocásticos define uma distribuição de probabilidade para várias redes, $P(G | \theta)$, na qual θ é um conjunto de parâmetros que governa as probabilidades das ligações e a estrutura de comunidades. Dado a escolha de θ , podemos gerar a instância de uma rede G . A inferência entra nesse processo de forma inversa, isto é, a partir de uma rede G , queremos identificar a escolha de θ que é mais provável de gerar G . Assim, ao encontrarmos θ , obtemos a separação da rede em seus módulos.

Existem diversas variações do SBM. Em especial, nesse trabalho, usamos uma variação desenvolvida por Peixoto [61]. Esse método é conhecido como modelo bayesiano de blocos estocásticos e permite encontrar estruturas modulares de forma que as redes geradas devem possuir exatamente a sequência de grau da rede. Essa formulação é mais abrangente e leva em conta uma maior complexidade da rede.

Entretanto, na sua forma mais simples, o modelo bayesiano de blocos estocásticos possui uma desvantagem quando aplicado a redes com muitos vértices. Nesse cenário, o modelo não é ideal para encontrar grupos relativamente pequenos. Para contornar esse problema, a abordagem conhecida como SBM aninhada, também desenvolvida por Peixoto, considera comunidades dentro de comunidades por meio de uma recursão realizada no SBM [62]. Com esse processo, o número máximo de grupos que pode ser inferido é maior e o modelo é mais ideal para identificar grupos pequenos em redes grandes. Uma descrição matemática detalhada desse modelo se encontra na referência [62]. Em nossas inferências da estrutura modular da rede de corrupção espanhola, utilizamos esse modelo mais complexo, conforme a implementação fornecida pelo pacote *graph-tool* [51].

Apêndice B: Conceitos de estatística

B.1 Teste de hipótese

O objetivo de um teste de hipótese [63] é determinar se podemos ou não rejeitar uma hipótese nula. A hipótese nula é definida, a princípio, ser verdadeira e afirma que a diferença entre duas medidas observadas é puramente aleatória; isto é, não existe diferença significativa entre as medidas.

B.1.1 Teste t de Welch

O teste t de Welch [64] verifica se as médias de duas amostras são significantemente diferentes. Nesse caso, a hipótese nula afirma que não existe diferença entre a média das amostras. Esse método é mais geral do que o teste de Student [65] e não assume amostras de tamanhos iguais, apenas variâncias similares. Nesse caso, definimos a estatística

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (\text{B.1})$$

com

$$s_p = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}}, \quad (\text{B.2})$$

na qual a amostra 1 possui média \bar{X}_1 , desvio padrão s_{X_1} e tamanho n_1 . Já a amostra 2 possui média \bar{X}_2 , desvio padrão s_{X_2} e tamanho n_2 .

A partir da distribuição de probabilidade obtida para a estatística t [64], o p -valor é simplesmente a soma das áreas abaixo da curva da distribuição de probabilidade maior do que $+t$ e menor do que $-t$. Isso representa a probabilidade da diferença das médias não estar no intervalo $\pm t$. Dessa forma, se escolhermos um intervalo de confiança de 95%, caso o p -valor seja menor ou maior do que 0.05, rejeitamos ou não a hipótese nula.

B.2 Medidas estatísticas

B.2.1 Média

No caso discreto, a média de um conjunto de N observações é dada por

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{N}. \quad (\text{B.3})$$

Já no caso contínuo, considerando uma densidade de probabilidade $f(x)$, a média da distribuição é calculada como

$$\langle x \rangle = \int_{-\infty}^{\infty} xf(x)dx. \quad (\text{B.4})$$

B.2.2 Desvio padrão

O desvio padrão mede o quanto disperso é um conjunto de dados em relação à sua média. No caso discreto, temos

$$\sigma = \sqrt{\frac{1}{N} \sum_n (x_n - \mu)^2}. \quad (\text{B.5})$$

Já no caso contínuo, considerando novamente uma densidade de probabilidade $f(x)$, o desvio padrão pode ser calculado como

$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x - \langle x \rangle)^2 f(x)dx}. \quad (\text{B.6})$$

B.2.3 Erro padrão da média

O erro padrão da média, por sua vez, representa o quanto representativa é a média obtida. O erro padrão da média informa o quanto discrepante é a média da amostra em relação a média da população. Essa medida é calculada como

$$SE = \frac{\sigma}{\sqrt{N}}. \quad (\text{B.7})$$

B.3 Distribuição acumulada

A distribuição acumulada de uma variável X representa a probabilidade de encontrar um valor x tal que $x < X$, ou ainda, $P(x \leq X)$. Como ilustração, considere uma distribuição exponencial $P(k)$ tal que

$$P(k) = \lambda e^{-k\lambda}. \quad (\text{B.8})$$

Nesse caso, o valor característico (ou valor médio) dessa distribuição pode ser calculado da seguinte maneira,

$$\langle k \rangle = \int_0^\infty k P(k) dk = \frac{1}{\lambda}. \quad (\text{B.9})$$

Além disso, sua distribuição acumulada pode ser escrita como

$$F(x) = P(X \leq x), \quad (\text{B.10})$$

ou seja,

$$F(x) = \int_0^x P(k) dk = 1 - e^{-x\lambda}. \quad (\text{B.11})$$

Outra forma comum de representar distribuições de probabilidade é por meio da distribuição acumulada complementar. No caso exponencial, temos

$$\overline{F(x)} = 1 - F(x) = e^{-x\lambda}. \quad (\text{B.12})$$

Calculando o logaritmo em ambos os lados da equação anterior, obtemos

$$\ln \overline{F(x)} = -x\lambda. \quad (\text{B.13})$$

Assim, em escala mono-logarítmica, essa equação se comporta como uma reta. Caso seja feita uma mudança de variável da forma $l = \frac{x}{\langle k \rangle}$, temos

$$\ln \overline{F(l\langle k \rangle)} = -l, \quad (\text{B.14})$$

ou seja, a distribuição acumulada complementar da exponencial nessa variável escalada é representada por uma reta com inclinação -1 em escala mono-logarítmica.

B.4 Método *bootstrap*

O método *bootstrap* [66] é um procedimento estatístico de reamostragem usado para construir intervalos de confiança. Por exemplo, se quisermos obter uma medida δ de uma amostra $X = \{x_1, x_2, \dots, x_n\}$, o intervalo de confiança dessa medida pode ser calculado da seguinte maneira.

A partir da amostra original X , escolhemos (de forma aleatória) k elementos para formar um novo conjunto X_i . Esse processo é realizado diversas vezes e calculamos a medida δ para cada conjunto X_i . Esse processo produz um conjunto de valores para a medida δ , ou seja,

$$\Delta = \{\delta(X_1), \delta(X_2), \dots, \delta(X_n)\}. \quad (\text{B.15})$$

Dessa forma, o intervalo de confiança da medida δ é calculado por meio de Δ . Se $Q_\beta[\Delta]$ representa o β -quantil da distribuição de probabilidade dos elementos de Δ , então um intervalo de confiança no nível α da medida δ é limitado inferiormente por $(Q_{\frac{\alpha}{2}}[\Delta])$ e superiormente por $(Q_{1-\frac{\alpha}{2}}[\Delta])$.

Referências Bibliográficas

- [1] Rotberg, R. I. *The Corruption Cure: How Citizens and Leaders Can Combat Graft* (Princeton University Press, Princeton, 2017).
- [2] What is corruption? – Corruptie. <http://www.corruptie.org/en/corruption/what-is-corruption/> (Acessado em: 09/10/2020).
- [3] Transparency International – What is corruption? <https://www.transparency.org/en/what-is-corruption> (Acessado em: 09/10/2020).
- [4] The costs of corruption: Values, economic development under assault, trillions lost, says Guterres. <https://news.un.org/en/story/2018/12/1027971> (Acessado em: 10/10/2020).
- [5] Mungiu-Pippidi, A. Corruption: Good governance powers innovation. *Nature News* **518**, 295, DOI: [10.1038/518295a](https://doi.org/10.1038/518295a) (2015).
- [6] Lapatinas, A., Litina, A. & Sartzetakis, E. S. Environmental projects in the presence of corruption. *International Tax and Public Finance* **26**, 103–144, DOI: [10.1007/s10797-018-9503-6](https://doi.org/10.1007/s10797-018-9503-6) (2019).
- [7] Ukwuije, C. B. *Corruption and Transformation in a Developing Economy* (AuthorHouse UK, Milton Keynes, 2019).
- [8] Uslaner, E. M. Corruption and Inequality. <https://core.ac.uk/download/pdf/6656858.pdf> (Acessado em: 10/10/2020).
- [9] López-Iturriaga, F. J. & Sanz, I. P. Predicting public corruption with neural networks: An analysis of spanish provinces. *Social Indicators Research* **140**, 975–998, DOI: [10.1007/s11205-017-1802-2](https://doi.org/10.1007/s11205-017-1802-2) (2018).
- [10] Public Opinion - European Commission. <https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/Survey/getSurveyDetail/search/corruption/surveyKy/2176> (Acessado em: 10/10/2020).

- [11] The costs of corruption across the European Union. <https://www.greens-efa.eu/en/article/document/the-costs-of-corruption-across-the-european-union/> (Acessado em: 10/10/2020).
- [12] Fifth evaluation round / Cinquième cycle d'évaluation. <https://www.coe.int/en/web/greco/evaluations/round-5-new> (Acessado em: 10/10/2020).
- [13] Transparency International España. <https://www.transparency.org/en/countries/spain> (Acessado em: 10/10/2020).
- [14] Arestis, P. & Sawyer, M. (eds.) *The Euro Crisis*. International Papers in Political Economy (Palgrave Macmillan UK, London, 2012).
- [15] Spanish unemployment hits record 5.64 million. <https://www.bbc.com/news/business-17866382> (Acessado em: 11/10/2020).
- [16] Bankruptcy epidemic slays Spain's businesses. <https://www.thelocal.es/20131206/spains-bankruptcy-epidemic-slays-giants-dwarfs> (Acessado em 10/11/2020).
- [17] Mungiu-Pippidi, A. The time has come for evidence-based anticorruption. *Nature Human Behaviour* **1**, 1–3, DOI: [10.1038/s41562-016-0011](https://doi.org/10.1038/s41562-016-0011) (2017).
- [18] Ribeiro, H. V., Alves, L. G. A., Martins, A. F., Lenzi, E. K. & Perc, M. The dynamical structure of political corruption networks. *Journal of Complex Networks* **6**, 989–1003, DOI: [10.1093/comnet/cny002](https://doi.org/10.1093/comnet/cny002) (2018).
- [19] Alves, L. G. A., Ribeiro, H. V. & Mendes, R. S. Scaling laws in the dynamics of crime growth rate. *Physica A: Statistical Mechanics and its Applications* **392**, 2672–2679, DOI: [10.1016/j.physa.2013.02.002](https://doi.org/10.1016/j.physa.2013.02.002) (2013).
- [20] Ribeiro, H. V., Hanley, Q. S. & Lewis, D. Unveiling relationships between crime and property in England and Wales via density scale-adjusted metrics and network tools. *PLOS ONE* **13**, e0192931, DOI: [10.1371/journal.pone.0192931](https://doi.org/10.1371/journal.pone.0192931) (2018).
- [21] da Cunha, B. R. & Gonçalves, S. Topology, robustness, and structural controllability of the Brazilian Federal Police criminal intelligence network. *Applied Network Science* **3**, 1–20, DOI: [10.1007/s41109-018-0092-1](https://doi.org/10.1007/s41109-018-0092-1) (2018).
- [22] Barabási, A.-L. & Pósfai, M. *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).
- [23] Stefan Thurner, Rudolf Hanel & Peter Klimek. *Introduction to the Theory of Complex Systems* (Oxford University Press, Oxford, 2018).

- [24] Scrapy | A Fast and Powerful Scraping and Web Crawling Framework. <https://scrapy.org/> (Último acesso em: 18/10/2020).
- [25] Welcome to Python.org. <https://www.python.org/> (Último acesso em: 18/10/2020).
- [26] IFES Election Guide | Elections: Spain Parliament. <https://www.electionguide.org/elections/id/3359/> (Último acesso em: 03/01/2021).
- [27] García, D. M. Davidmogar/genderator. <https://github.com/davidmogar/genderator> (2020).
- [28] INE. Instituto Nacional de Estadística. <https://www.ine.es/> (Último acesso em: 05/11/2020).
- [29] ROWE, D. C., VAZSONYI, A. T. & FLANNERY, D. J. Sex differences in crime: Do means and within-sex variation have similar causes. *Journal of Research in Crime and Delinquency* **32**, 84–100, DOI: [10.1177/0022427895032001004](https://doi.org/10.1177/0022427895032001004) (2016).
- [30] Walker, J. & Maddan, S. *Understanding Statistics For The Social Sciences, Criminal Justice, And Criminology* (Jones & Bartlett Learning, Burlington, 2012).
- [31] Bennett, S., Farrington, D. P. & Huesmann, L. R. Explaining gender differences in crime and violence: The importance of social cognitive skills. *Aggression and Violent Behavior* **10**, 263–288, DOI: [10.1016/j.avb.2004.07.001](https://doi.org/10.1016/j.avb.2004.07.001) (2005).
- [32] Myung, I. J. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* **47**, 90–100, DOI: [10.1016/S0022-2496\(02\)00028-7](https://doi.org/10.1016/S0022-2496(02)00028-7) (2003).
- [33] Baker, W. E. & Faulkner, R. R. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review* **58**, 837–860, DOI: [10.2307/2095954](https://doi.org/10.2307/2095954) (1993).
- [34] Chami, G. F., Ahnert, S. E., Kabatereine, N. B. & Tukahebwa, E. M. Social network fragmentation and community health. *Proceedings of the National Academy of Sciences* **114**, E7425–E7431, DOI: [10.1073/pnas.1700166114](https://doi.org/10.1073/pnas.1700166114) (2017).
- [35] Braunstein, A., Dall'Asta, L., Semerjian, G. & Zdeborová, L. Network dismantling. *Proceedings of the National Academy of Sciences* **113**, 12368–12373, DOI: [10.1073/pnas.1605083113](https://doi.org/10.1073/pnas.1605083113) (2016).
- [36] da Cunha, B. R., González-Avella, J. C. & Gonçalves, S. Fast fragmentation of networks using module-based attacks. *PLOS ONE* **10**, e0142824, DOI: [10.1371/journal.pone.0142824](https://doi.org/10.1371/journal.pone.0142824) (2015).

- [37] Ren, X.-L., Gleinig, N., Helbing, D. & Antulov-Fantulin, N. Generalized network dismantling. *Proceedings of the National Academy of Sciences* **116**, 6554–6559, DOI: [10.1073/pnas.1806108116](https://doi.org/10.1073/pnas.1806108116) (2019).
- [38] Cavallaro, L. *et al.* Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia. *PLOS ONE* **15**, e0236476, DOI: [10.1371/journal.pone.0236476](https://doi.org/10.1371/journal.pone.0236476) (2020).
- [39] Krebs, V. E. Mapping Networks of Terrorist Cells. *Connections* **24**, 43 (2002).
- [40] Wood, G. The structure and vulnerability of a drug trafficking collaboration network. *Social Networks* **48**, 1–9, DOI: [10.1016/j.socnet.2016.07.001](https://doi.org/10.1016/j.socnet.2016.07.001) (2017).
- [41] Bunde, A. & Havlin, S. (eds.) *Fractals and Disordered Systems* (Springer-Verlag, Berlin, 1996).
- [42] País, E. Un juez investiga el despacho que fundó Montoro tras una denuncia de Anticorrupción. *El País* (2017). https://elpais.com/politica/2017/06/27/actualidad/1498561386_471191.html (Último acceso em: 04/02/2021).
- [43] La amnistía fiscal del PP y sus chapuzas | España Home | EL MUNDO. <https://www.elmundo.es/espana/2017/06/12/593d5f03e2704ef06a8b4631.html> (Último acesso em: 04/02/2021).
- [44] Mestre, J. F. El tribunal del caso Over confirma la financiación ilegal del PP de Baleares. <https://www.diariodemallorca.es/mallorca/2019/06/15/tribunal-caso-over-confirma-financiacion-2867112.html> (Último acesso em: 04/01/2021).
- [45] Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7821–7826, DOI: [10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799) (2002).
- [46] Cherifi, H., Palla, G., Szymanski, B. K. & Lu, X. On community structure in complex networks: Challenges and opportunities. *Applied Network Science* **4**, 1–35, DOI: [10.1007/s41109-019-0238-9](https://doi.org/10.1007/s41109-019-0238-9) (2019).
- [47] Ibrahim, N. M. A. & Chen, L. Link prediction in dynamic social networks by integrating different types of information. *Applied Intelligence* **42**, 738–750, DOI: [10.1007/s10489-014-0631-0](https://doi.org/10.1007/s10489-014-0631-0) (2015).
- [48] Stegehuis, C., van der Hofstad, R. & van Leeuwaarden, J. S. H. Epidemic spreading on complex networks with community structures. *Scientific Reports* **6**, 29748, DOI: [10.1038/srep29748](https://doi.org/10.1038/srep29748) (2016).

- [49] Weng, L., Menczer, F. & Ahn, Y.-Y. Virality prediction and community structure in social networks. *Scientific Reports* **3**, 2522, DOI: [10.1038/srep02522](https://doi.org/10.1038/srep02522) (2013).
- [50] Calderoni, F., Brunetto, D. & Piccardi, C. Communities in criminal networks: A case study. *Social Networks* **48**, 116–125, DOI: [10.1016/j.socnet.2016.08.003](https://doi.org/10.1016/j.socnet.2016.08.003) (2017).
- [51] Graph-tool: Efficient network analysis with Python. <https://graph-tool.skewed.de/> (Último acesso em: 29/01/2021).
- [52] Holten, D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* **12**, 741–748, DOI: [10.1109/TVCG.2006.147](https://doi.org/10.1109/TVCG.2006.147) (2006).
- [53] Weisstein, E. W. Least Squares Fitting. <https://mathworld.wolfram.com/LeastSquaresFitting.html> (Último acesso em: 07/02/2021).
- [54] Scipy.stats.linregress — SciPy v1.6.0 Reference Guide. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html> (Último acesso em: 07/02/2021).
- [55] Newman, M. E. J. Mixing patterns in networks. *Physical Review E* **67**, 026126, DOI: [10.1103/PhysRevE.67.026126](https://doi.org/10.1103/PhysRevE.67.026126) (2003).
- [56] Peixoto, T. P. Centrality measures. <https://graph-tool.skewed.de/static/doc/centrality.html> (Último acesso em: 23/11/2020).
- [57] Erdős, P. & Rényi, A. *On the Evolution of Random Graphs* (Princeton University Press, Princeton, 2011).
- [58] Fortunato, S. Community detection in graphs. *Physics Reports* **486**, 75–174, DOI: [10.1016/j.physrep.2009.11.002](https://doi.org/10.1016/j.physrep.2009.11.002) (2010).
- [59] Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137, DOI: [10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7) (1983).
- [60] Jebara, T. *Machine Learning: Discriminative and Generative*. The Springer International Series in Engineering and Computer Science (Springer, New York, 2004).
- [61] Peixoto, T. P. Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E* **95**, 012317, DOI: [10.1103/PhysRevE.95.012317](https://doi.org/10.1103/PhysRevE.95.012317) (2017). [1610.02703](https://arxiv.org/abs/1610.02703).
- [62] Peixoto, T. P. Bayesian stochastic blockmodeling. *arXiv [cond-mat, physics:physics, stat]* 289–332, DOI: [10.1103/9781119483298.ch11](https://doi.org/10.1103/9781119483298.ch11) (2019).

- [63] Statistics - Hypothesis testing. <https://www.britannica.com/science/statistics/Hypothesis-testing> (Último acesso em: 28/11/2020).
- [64] Welch, B. L. The generalization of Student's problem when several different population variances are involved. *Biometrika* **34**, 28–35, DOI: [10.1093/biomet/34.1-2.28](https://doi.org/10.1093/biomet/34.1-2.28) (1947).
- [65] Student's t-test | Definition, Formula, & Example. <https://www.britannica.com/science/Students-t-test> (Último acesso em: 28/11/2020).
- [66] Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman and Hall/CRC, London, 1994).