

UFRRJ
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
MATEMÁTICA E COMPUTACIONAL

DISSERTAÇÃO

**Equilíbrio estrutural aplicado à detecção de casos de
corrupção**

Vitor dos Santos Ponciano

Seropédica, RJ
Fevereiro, 2017



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXATAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E
COMPUTACIONAL**

**EQUILÍBRIO ESTRUTURAL APLICADO À DETECÇÃO DE CASOS
DE CORRUPÇÃO**

VITOR DOS SANTOS PONCIANO

Sob a orientação do professor

Ronaldo Malheiros Gregório

Sob a Co-orientação da professora

Rosa Maria Videira de Figueiredo

Dissertação submetida como requisito parcial para a obtenção do grau de Mestre em Ciências, no Curso de Pós-Graduação em Modelagem Matemática e Computacional, área de concentração em Modelagem Matemática e Computacional.

Seropédica, RJ

Fevereiro, 2017

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

P792e Ponciano, Vitor dos Santos, 1989-
Equilíbrio estrutural aplicado à detecção de casos de
corrupção / Vitor dos Santos Ponciano. - 2017.
57 f.: il.

Orientador: Ronado Malheiros Gregório.
Coorientadora: Rosa Maria Videira de Figueiredo.
Dissertação(Mestrado). -- Universidade Federal
Rural do Rio de Janeiro, Progrma de Pós-Graduação em
Modelagem Matemática e Computacional, 2017.

1. Equilíbrio estrutural. 2. Grafos de Sinais. 3.
Detecção de Corrupção. I. Gregório, Ronado Malheiros,
1978-, orient. II. Figueiredo, Rosa Maria Videira de,
1972-, coorient. III Universidade Federal Rural do
Rio de Janeiro. Progrma de Pós-Graduação em Modelagem
Matemática e Computacional. IV. Título.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXATAS
MODELAGEM MATEMÁTICA E COMPUTACIONAL

Vitor dos Santos Ponciano

Dissertação submetida como requisito parcial para a obtenção do grau de Mestre em Ciências, no Curso de Pós-Graduação em Modelagem Matemática e Computacional, área de concentração em Modelagem Matemática e Computacional.

Dissertação aprovada em 21 de Fevereiro de 2017

Ronaldo Malheiros Gregório, D.Sc.
UFRRJ
(Orientador)

Marcelo Dib Cruz, D.Sc.
UFRRJ

Yuri Abitbol de Menezes Frota, D.SC.
UFF

Seropédica, RJ
Fevereiro, 2017

Agradecimentos

Agradeço primeiramente a Deus por me conceder forças para prosseguir nesta jornada. A minha família por estar sempre do meu lado, principalmente nos momentos ruins. Aos amigos, aos meu orientadores Ronaldo Grégorio e Rosa Maria Figueiredo pela paciência e ajuda durante todo processo de construção deste trabalho, à CAPES pelo apoio financeiro, à LIA - CERI (Université d'Avignon, França) por conceder o intercambio e ferramentas para a realização dos testes computacionais e a todos aqueles que contribuíram de forma significativa para o desenvolvimento deste trabalho.

“A matemática, vista corretamente, possui não apenas verdade, mas também suprema beleza -
uma beleza fria e austera, como a da escultura.”
(Bertrand Russell)

“O homem erudito é um descobridor de fatos que já existem - mas o homem sábio é um criador
de valores que não existem e que ele faz existir.”
(Albert Einstein)

Resumo

Em 1946, Heider desenvolveu a Teoria de Grafos de Sinais a fim de descrever as relações emocionais entre as pessoas pertencentes ao mesmo grupo social. Neste trabalho, estudamos problemas de particionamento de grafo associados com equilíbrio estrutural, que na literatura de Ciência da Computação são conhecidos como problemas de correlação de partições ou, em inglês, *correlation clustering* (CC), além de uma versão relaxada (RCC). As soluções dos problemas CC e RCC foram anteriormente utilizadas na literatura como ferramentas para a avaliação de equilíbrio estrutural numa rede social. O objetivo deste trabalho é aplicar as soluções destes problemas na detecção de corrupção em contratos públicos. Para esse fim, são utilizadas formulações de programação linear inteira existentes na literatura para estes problemas. Além disso, é discutido também um modelo probabilístico para o Equilíbrio Estrutural e as soluções destes problemas são aplicadas à detecção de casos de corrupções existentes em licitações públicas.

Palavras-chaves: Equilíbrio estrutural, Grafos de sinais, Correlation Clustering, Detecção de Corrupção.

Abstract

In 1946, Heider developed Signal Graph Theory with the purpose of describing the emotional relationships between people pertaining to the same social group. In this work, we study graph partitioning problems associated with structural balance. These problems are known in the computer science literature as partition correlation problems: correlation clustering (CC) and a relaxed version (RCC). The solution of CC and RCC problems has been previously used in the literature as a tool for the evaluation of structural balance in a social network. The aim of this work is to apply the solution of these problems in the detection of corruption in public contracts. We describe integer linear programming formulations from the literature for these problems. We also discuss a probabilistic model for Structural balance and the solution of these problems applied to the detection of cases of corruption existing in public bids.

Key-words: Structural balance, Signal Graphs, Correlation Clustering, Corruption Detection.

Lista de ilustrações

Figura 1 – Triângulos equilibrados	14
Figura 2 – Triângulos desequilibrados	14
Figura 3 – Semi-ciclo em grafo de sinais	15
Figura 4 – Primeira Guerra Mundial. Fonte (ESTRADA; BENZI, 2014)	16
Figura 5 – Grafo equilibrado	18
Figura 6 – Grafo desequilibrado	18
Figura 7 – A diferença entre CC e RCC. Fonte (LEVORATO, 2015)	20
Figura 8 – Exemplo da prova do teorema 2.3.1. Fonte (FIGUEIREDO; MOURA, 2013)	22
Figura 9 – Grafo sinalizado G com $n = 3$	29
Figura 10 – Estado livre de corrupção	36
Figura 11 – Corrupção sem <i>state capture</i>	37
Figura 12 – Estado parcialmente apropriado	37
Figura 13 – Estado completamente capturado	38
Figura 14 – Run1	43
Figura 15 – Run1[P1]	44
Figura 16 – Run1[P2]	45
Figura 17 – Run1[P3]	46
Figura 18 – Run1[P4]	47
Figura 19 – Run1[P5]	48
Figura 20 – Run1[P6]	49
Figura 21 – Run1[P7]	50
Figura 22 – Run1[P8]	51
Figura 23 – Run1[P9]	52
Figura 24 – Run1[P10]	53

Lista de tabelas

Tabela 1 – Literatura Instâncias	26
Tabela 2 – Instância MonkT3	26
Tabela 3 – Instância MonkT4	26
Tabela 4 – Instância House B	27
Tabela 5 – Instância House C	27
Tabela 6 – Instância Mickney	28
Tabela 7 – Instância Newcomb	28
Tabela 8 – Tabelas de probabilidade	39
Tabela 9 – Resultados CC e CC probabilístico	41
Tabela 10 – Resultados RCC e RCC probabilístico	42
Tabela 11 – Resultados CC e CC probabilístico Run1	43

Sumário

1	Introdução	12
1.1	Objetivos	13
1.2	Estrutura	13
2	Equilíbrio em grafo de sinais	14
2.1	Notações e definições de Equilíbrio Estrutural	14
2.1.1	Grafos de Sinais	14
2.1.2	Equilíbrio em Grafos de Sinais	14
2.2	Modelo de PLI para Correlação de Clusters	19
2.3	Equilíbrio Estrutural Relaxado	19
2.4	Modelo de PLI para o problema RCC	22
2.4.1	Uma formulação por representantes	22
2.4.2	Uma formulação usando divisão por clusters	24
2.4.3	Experimentos Computacionais com as formulações do RCC	25
2.4.3.1	Instâncias Monastério Sampson	26
2.4.3.2	Instâncias House	27
2.4.3.3	Instância Mickney	27
2.4.3.4	Instância Newcomb	28
3	Grafos de sinais probabilístico	29
3.1	Caso probabilístico para CC	29
3.1.1	Modelo de Correlação de Clustering Probabilístico	31
3.2	Caso probabilístico para o RCC	32
3.2.1	Modelo Relaxado de Correlação de Clustering Probabilístico	34
4	Análise de Corrupção	35
4.1	O Índice de Risco de Corrupção (IRC) em <i>state capture</i>	35
4.2	Configurações das redes de corrupção em grafos de sinais	36
4.3	Experimentos Computacionais em Redes de Corrupção	38
4.3.1	Estudo das Instâncias	39
4.3.2	Análise das instâncias	40
4.3.2.1	Análise Run1	40
5	Trabalhos futuros e conclusão	54
	Referências	55

1 Introdução

A teoria do Equilíbrio Estrutural tem sido uma preocupação para muitos psicólogos sociais e sociólogos. O estudo sobre Equilíbrio estrutural iniciou-se com (HEIDER, 1946)¹ na psicologia social para modelar situações onde há relações positivas, que representam amizade, e negativas, que representam inimizades, entre pessoas. A Teoria de Equilíbrio Estrutural foi formalizada por (CARTWRIGHT; HARARY, 1967) e estabelece que um grupo social equilibrado pode ser dividido em dois subgrupos mutuamente hostis, cada qual apresentando solidariedade interna. Em diversos trabalhos da literatura, o equilíbrio estrutural é aplicado a redes sociais, como por exemplo (ADEJUMO; DUIMERING; ZHONG, 2008), (ABELL; LUDWIG, 2009), (DOREIAN; MRVAR, 1996), (ESTRADA; BENZI, 2014), (MENDONCA et al., 2015).

O termo *Clustering* é conhecido em Ciência da Computação como a atividade de particionar elementos individuais em grupos com base em sua similaridade. Como ressaltado em (FROTA Y., 2010) e (DRUMMOND L.; LEVORATO, 2013), podemos identificar problemas de clustering definidos em grafos de sinais em diferentes áreas científicas. Isso pode ser constatado em (BANSAL; BLUM; CHAWLA, 2002), (DASGUPTA et al., 2007), (TRAAG; BRUGGEMAN, 2009), (HUFFNER; BETZLER; NIEDERMEIER, 2007), (FIGUEIREDO; FROTA, 2014). Medidas alternativas para o equilíbrio estrutural em problemas de clustering também foram discutidas recentemente em (DOREIAN; MRVAR, 2009), (FIGUEIREDO; MOURA, 2013), (DOREIAN; MRVAR, 2016).

O conceito de *State capture* foi introduzido por pesquisadores do Banco Mundial para descrever a situação em alguns países da antiga esfera soviética, na Europa Oriental, e ele é definido como o conjunto de esforços que determinadas empresas dedicam para influenciar na formação das leis, das políticas e dos regulamentos do Estado em sua própria vantagem, como por exemplo pagamentos ilícitos a oficiais públicos (HELLMAN et al., 2000). As empresas foram deliberadamente adotando essas estratégias para influenciar políticos e funcionários, com vista à promoção de mudanças legislativas, obtendo assim favorecimentos.

O estudo de clustering para a compreensão do conceito de state capture foi introduzido nos trabalhos (FAZEKAS; TÓTH, 2014a) e (FAZEKAS; TÓTH; KING, 2013). Tais pesquisadores desenvolveram um novo quadro conceitual e analítico para medir o *state capture* com base em redes de contratos públicos. Para este fim, eles primeiro estabeleceram uma medida robusta de riscos de corrupção nas transações de contratos públicos, focando nas relações entre pares de emissores e fornecedores, além do Índice de Risco de Corrupção (IRC).

A corrupção neste ponto de vista é notada nas mudanças políticas entre comandos governamentais (por exemplo, mudança de prefeitos, governadores, presidentes, etc.), momento onde se torna vulnerável a administração de bens públicos básicos, e nas mudanças de comando em grupos de poderosas empresas (predominantemente públicas), que passam ao *status* de recém privatizadas e que são líderes em suas respectivas áreas de atuação. Para competir nesse entorno, as novas empresas adotam a estratégia de *state capture* afim de assegurar direitos às propriedades junto à funcionários públicos e incorporar novas vantagens no novo mercado econômico e

¹ Fritz Heider (1896 - 1988) foi um psicólogo austríaco-americano cujo trabalho relaciona-se à psicologia das relações interpessoais, teoria base para grande parte dos estudos da psicologia social contemporânea. Ele defendia que os sentimentos positivos e negativos precisam ser representados de forma a minimizar os sentimentos conflitantes e maximizar um estado afetivo (HARVEY, 1989).

político.

Usualmente, grafos de sinais são tratados como grafos precisos, onde a cada aresta é associado um sinal positivo ou negativo. Contudo, obter informações sobre sinais de arestas nem sempre é fácil na prática. Em inúmeras aplicações, associamos a cada aresta um peso (normalmente, valores entre 0 e 1) para representar a certeza ou a confiança no sinal da relação. Uma vez definido o grafo de sinais e os pesos das arestas, a estratégia adotada é resolver uma versão ponderada do problema de particionamento de grafos. Neste trabalho propomos que os pesos das arestas sejam tratados como probabilidades e uma versão do problema, onde procuramos a partição do grafo que minimiza o desequilíbrio esperado, pois no estudo da corrupção as relações são mais propensas a apresentar uma natureza incerta.

1.1 Objetivos

Dentre as metas propostas para esse trabalho, podemos destacar a:

- descrição de uma versão probabilística para o CC e o RCC, isto é, uma formulação para problemas de particionamento de grafo CC e RCC, com pesos para as arestas baseados em probabilidades, onde buscamos a partição do grafo que minimiza o desequilíbrio esperado.
- utilização da solução da versão probabilística como ferramenta na detecção de casos de corrupção em licitações públicas.

1.2 Estrutura

No segundo capítulo, apresentamos a definição de grafos de sinais e equilíbrio estrutural, descrevemos duas formulações para o problema RCC e realizamos testes computacionais com instâncias da literatura.

No terceiro capítulo, formulamos uma versão probabilística para CC e o RCC, onde buscamos a partição do grafo que minimiza o desequilíbrio esperado.

No quarto capítulo, detalhamos a noção de *state capture*, índice de corrupção e realizamos experimentos computacionais com instâncias de redes de corrupção.

2 Equilíbrio em grafo de sinais

Neste capítulo definiremos o conceito de grafos de sinais equilibrados e mostraremos os modelos de programação inteira para os problemas de particionamento de grafos associadas a este conceito.

2.1 Notações e definições de Equilíbrio Estrutural

2.1.1 Grafos de Sinais

Definição 2.1.1. *Seja $G = (V, A)$ um grafo sem laços e não direcionado, onde V é um conjunto de n vértices e A é o conjunto de aresta. Considere agora uma função $s = A \rightarrow \{+, -\}$ que atribui um sinal a cada arco ou aresta de A . $G = (V, A)$, juntamente com s , denotado aqui por $G = (V, A, s)$, é denominado grafo de sinais. Além disso, uma aresta $a \in A$ é dita negativa se $s(a) = -$ e positiva se $s(a) = +$. Os conjuntos A^+ e A^- , são respectivamente, o conjunto das arestas positivas e negativas em um grafo de sinais. Note que $A = A^+ \cup A^-$.*

2.1.2 Equilíbrio em Grafos de Sinais

A primeira definição de equilíbrio dada por (HEIDER, 1946) baseia-se na ideia de um triângulo, cujo os vértices são dois atores, p e q , e algum objeto, x .

Definição 2.1.2. *Um triângulo é dito equilibrado se o produto dos sinais que ele contém é positivo e desequilibrado quando o produto de seus sinais é negativo.*

Vemos nas Figuras 1 e 2 as ilustrações desta definição, onde arestas positivas e negativas são representadas, respectivamente, pelas cores verde e linhas vermelhas quebradas.

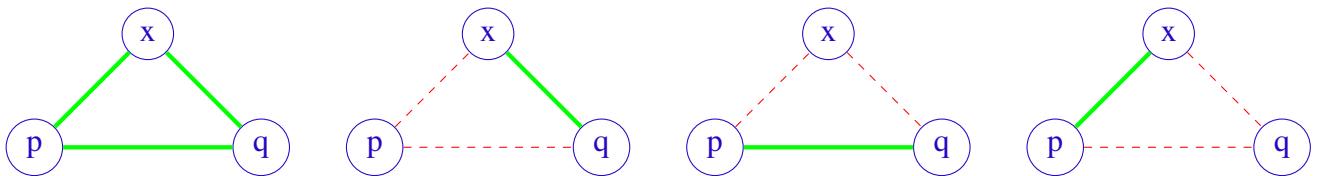


Figura 1: Triângulos equilibrados

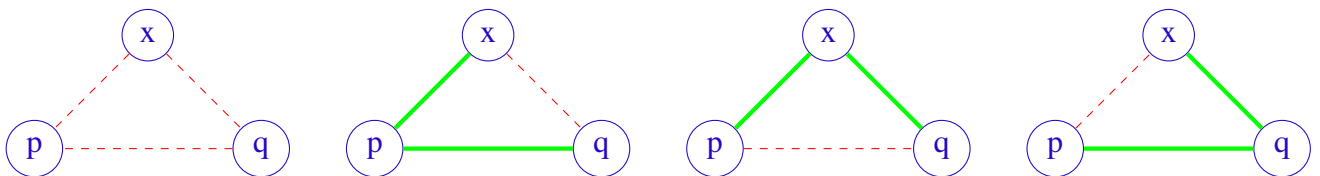


Figura 2: Triângulos desequilibrados

(CARTWRIGHT; HARARY, 1967) formalizaram a teoria de Heider afirmando que um grupo social é equilibrado quando pode ser dividido em dois subgrupos mutuamente antagonista, cada qual apresentando solidariedade interna. (DAVIS, 1967) estendeu esta definição para casos que englobam mais de dois subgrupos antagonistas.

Definição 2.1.3. Sejam $a_i = u_i u_{i+1} \in G, i \in \{1, \dots, k\}$, arestas de G , para algum inteiro positivo k . A sequência $W = a_1 a_2 \dots a_k$ é dita um passeio do vértice u_1 ao vértice u_{k+1} , de tamanho k . Assume-se portanto que a_i é adjacente a a_{i+1} para todo $i \in \{1, \dots, k\}$. Se a direção das linhas não for levada em consideração então temos o que denominamos por semi-passeio.

Definição 2.1.4. Seja $W = a_1 a_2 \dots a_k$ um passeio (semi-passeio).

- (i) W é dito fechado, se $u_1 = u_{k+1}$;
- (ii) W é denominado ciclo, se W é fechado e $u_i \neq u_j$, para todo $i \neq j$, com $i, j \in \{1, \dots, k\}$.
- (iii) W é um semi-ciclo se W é um ciclo em um semi-passeio.

Definição 2.1.5. Um ciclo (semi-ciclo) é positivo se o produto dos sinais de suas arestas (arcos) é positivo e negativo se o produto é negativo.

Definição 2.1.6. Um grafo de sinais $G = (V, A, s)$ é equilibrado se todos os seus ciclos (semi-ciclos) são positivos.

Exemplo 2.1.1. Considerando o semi-ciclo negativo $W = (u_1 u_2)(u_2 u_3)(u_3 u_4)(u_4 u_1)$ no grafo de sinais definido na figura 3, temos, pela definição 2.1.6, que ele é desequilibrado.

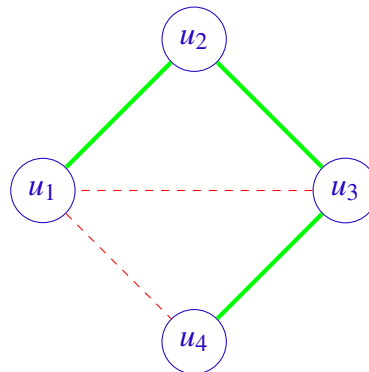


Figura 3: Semi-ciclo em grafo de sinais

Segundo (DOREIAN; MRVAR, 1996) as primeiras teorias de psicologias sociais ¹ postulam uma tendência de solução para estruturas equilibradas. Isso significa que, para verificar empiricamente, se uma rede de sinais está mudando com tempo em direção ao equilíbrio (ou não), torna-se necessário medir até que ponto essa rede está desequilibrada, a fim de testar a hipótese teórica de movimento em direção a estruturas equilibradas.

Exemplo 2.1.2. Na Figura 4, podemos ver como o exemplo de Equilíbrio Estrutural ocorre na evolução de uma rede social que está mudando, em relação ao tempo, em direção ao equilíbrio. Os grafos de sinais, representam a análise de evolução das relações entre os principais atores da Primeira Guerra Mundial. Ver (ESTRADA; BENZI, 2014). Estes correspondem a evolução do equilíbrio global entre os seis principais atores da Primeira Guerra Mundial em diferentes épocas. Sólidas linhas pretas representam alianças e linhas vermelhas quebradas representam conflitos (GB representa Gra-Bretanha, Ru: Rússia, Ge: Alemanha, Fr: França, AH: Império Austro-Húngaro e It: Itália).

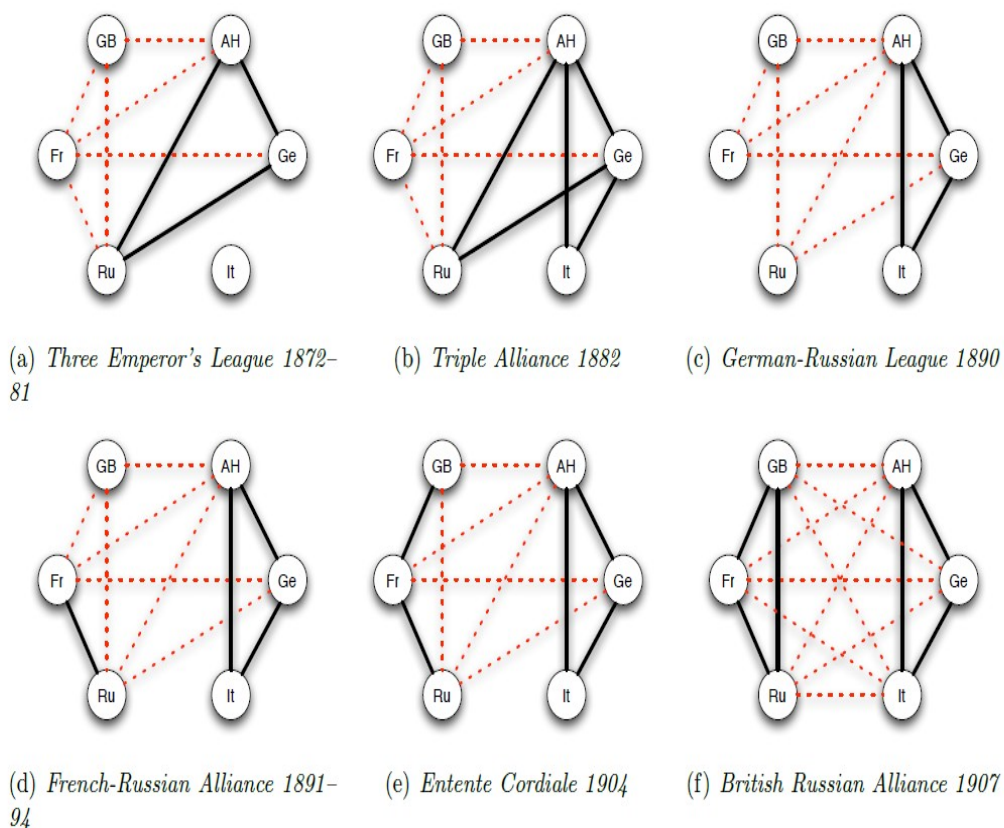


Figura 4: Primeira Guerra Mundial. Fonte (ESTRADA; BENZI, 2014)

A história da Figura 4 começa com a Liga dos Três Imperadores (1872 a 1881) que uniu Alemanha, Áustria-Hungria, e Rússia. A Tríplice Aliança foi formada em 1882, que juntou à Alemanha, à Áustria-Hungria e Itália em um bloco que continuou até a Primeira Guerra Mundial. Em 1890, um acordo bipartido entre a Alemanha e a Rússia perdeu seu valor legal e isso levou à criação de um governo francês-Russo durante o período 1891-94. Posteriormente, um Entente Cordiale entre a França e a Grã-Bretanha foi consumado em 1904, e então um acordo britânico-russo em 1907, que ligou França, Grã-Bretanha e Rússia na Triple Entente e culminou na Primeira Guerra Mundial.

A seguir, apresentamos uma série de definições necessárias para estabelecermos as medidas de equilíbrio usadas ao longo deste trabalho.

¹ Psicologia social é um ramo da psicologia que estuda como as pessoas pensam, influenciam e se relacionam umas com as outras (ALVARO; GARRIDO, 2006).

Definição 2.1.7. Seja $G = (V, A, s)$ um grafo de sinais sem laços e não direcionado, onde V é um conjunto formado por n vértices. Uma partição de V é qualquer coleção P de subconjuntos não vazios, tais que todo vértice de V pertence a exatamente um dos elementos de P . Neste texto uma partição de V em l subconjuntos é denominada uma l -partição.

Definição 2.1.8. Para um subconjunto de vértices $S \subseteq V$, defina $A[S] = \{(i, j) \in A \mid i, j \in S\}$. $A[S]$ é denominado subconjunto de arestas induzidas por S . Para dois subconjuntos de vértice S e H , com $S \subseteq V$ e $H \subseteq V$, definimos $A[S, H] = \{(i, j) \in A \mid i \in S \text{ e } j \in H\}$. Note que $A[S, S] = A[S]$.

Definição 2.1.9. Considere um parâmetro k inteiro, com $1 \leq k \leq n$. Um grafo de sinais $G = (V, A, s)$ é k -equilibrado se podemos dividir os vértices de V em conjuntos $S_1, S_2, \dots, S_k$, dois a dois disjuntos, tais que

$$\bigcup_{1 \leq i \leq k} A[S_i] = A^+. \quad (2.1)$$

Equivalentemente, um grafo de sinais $G = (V, A, s)$ é dito k -equilibrado se existe uma l -partição $P = \{S_1, \dots, S_l\}$ de V , com $l \leq k$, tal que

$$\bigcup_{1 \leq i \leq l} A[S_i] = A^+ \quad (2.2)$$

Definição 2.1.10. Considere uma l -partição de V , $P = \{S_1, S_2, \dots, S_l\}$. As arestas externas e as arestas internas relacionadas à essa partição são definidas respectivamente por

$$\bigcup_{1 \leq i \neq j \leq l} A[S_i : S_j] \quad (2.3)$$

e

$$\bigcup_{1 \leq i \leq l} A[S_i] \quad (2.4)$$

Definição 2.1.11. Seja w_a um peso positivo associado a uma aresta $a \in A$. Além disso, para $1 \leq i \neq j \leq l$ sejam

$$\Omega^+(S_i, S_j) = \sum_{a \in A^+ \cap A[S_i : S_j]} w_a \quad (2.5)$$

e

$$\Omega^-(S_i, S_j) = \sum_{a \in A^- \cap A[S_i : S_j]} w_a. \quad (2.6)$$

Segundo a definição de Equilíbrio Estrutural, o desequilíbrio $I(P)$ de uma partição P é definido como o somatório dos pesos das arestas internas negativas mais o somatório dos pesos das arestas externas positivas. Isto é,

$$I(P) = \sum_{1 \leq i \leq l} \Omega^-(S_i, S_i) + \sum_{1 \leq i \neq j \leq l} \Omega^+(S_i, S_j). \quad (2.7)$$

Exemplo 2.1.3. Temos dois exemplos descritos nas Figuras 5 e 6. Vamos utilizar a mesma partição $P = \{\{1, 2, 3, 6\}, \{4, 5\}\}$. Note que um grafo é equilibrado com a partição P e o outro, desequilibrado.

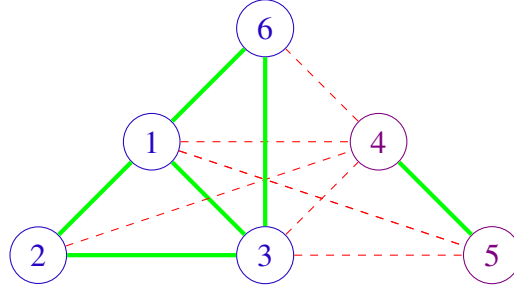


Figura 5: Grafo equilibrado

1. $P = \{\{1, 2, 3, 6\}, \{4, 5\}\}$,
2. $A^+ = \{(1, 2), (1, 3), (1, 6), (2, 3), (3, 6), (4, 5)\}$,
3. $S1 = \{1, 2, 3, 6\}$ e $S2 = \{4, 5\}$,
4. $A[S1] = \{(1, 2), (1, 3), (1, 6), (2, 3), (3, 6)\}$,
5. $A[S2] = \{(4, 5)\}$,
6. $A[S1] \cup A[S2] = A^+$, logo o grafo da figura 4 é equilibrado.

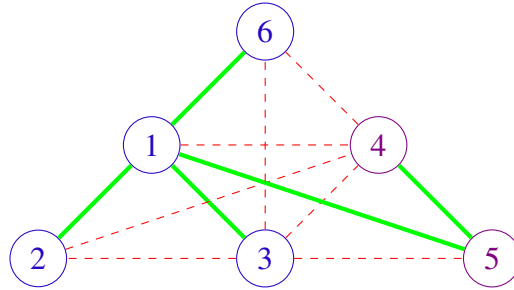


Figura 6: Grafo desequilibrado

1. $P = \{\{1, 2, 3, 6\}, \{4, 5\}\}$,
2. $A^+ = \{(1, 2), (1, 3), (1, 6), (1, 5), (4, 5)\}$,
3. $S1 = \{1, 2, 3, 6\}$ e $S2 = \{4, 5\}$,
4. $A[S1] = \{(1, 2), (1, 3), (3, 6), (1, 6)\}$,
5. $A[S2] = \{(4, 5)\}$,
6. $A[S1] \cup A[S2] \neq A^+$. Logo, grafo da figura 5 é desequilibrado.

Finalmente, estamos aptos a definir formalmente o problema de Correlação de Clusters.

(Problema CC) 1. *Seja $G = (V, A, s)$ um grafo de sinais e w_a um peso positivo associado a cada aresta $a \in A$. O problema de Correlação de Clusters ou, em inglês, Correlation Clustering (CC), consiste em determinar uma partição P de V , tal que o desequilíbrio $I(P)$ é mínimo.*

O problema CC é NP-completo. Isso pode ser visto em (BANSAL; BLUM; CHAWLA, 2002). Ao contrário de outros algoritmos de clustering, o CC não requer a escolha dos k -clusters antecipadamente, porque a meta de minimizar o desequilíbrio independe do número de clusters.

2.2 Modelo de PLI para Correlação de Clusters

Como ressaltado em (FIGUEIREDO; MOURA, 2013), a formulação clássica para o problema CC consiste em um modelo de programação linear inteira (PLI) proposto para os problemas de clustering não-capacitados (TRICK, 1996). Nesta formulação, uma variável de decisão binária x_{ij} é atribuída a cada par de vértices $i, j \in V, i \neq j$, e definida da seguinte maneira: $x_{ij} = 0$ se i e j estão no mesmo cluster e $x_{ij} = 1$, caso contrário. O modelo minimiza o desequilíbrio total.

$$\text{Minimizar} \quad \sum_{(i,j) \in A^-} w_{ij}(1 - x_{ij}) + \sum_{(i,j) \in A^+} w_{ij}x_{ij} \quad (2.8)$$

$$x_{ip} + x_{pj} \geq x_{ij}, \quad \forall i, p, j \in V, \quad (2.9)$$

$$x_{ij} = x_{ji}, \quad \forall i, j \in V, \quad (2.10)$$

$$x_{ij} \in \{0, 1\}. \quad (2.11)$$

As desigualdades triangulares representadas pela equação (2.9) dizem que, se i e p estão no mesmo cluster, assim como p e j , então os vértices i e j devem estar no mesmo cluster. O conjunto de restrições (2.10), escritas para todo $i, j \in V$, estabelece que as variáveis x_{ij} e x_{ji} sempre assumem o mesmo valor nesta formulação. As restrições em (2.11) indicam que as variáveis são binárias, enquanto a função objetivo (2.8) minimiza o desequilíbrio, definido em (2.7).

2.3 Equilíbrio Estrutural Relaxado

Em (DOREIAN, 2008), a definição para grafos de sinais k -equilibrados foi informalmente estendida para incluir processos relevantes (polarização, mediação, popularidade diferenciada e subgrupo com hostilidade interna) que originalmente eram vistos como violações do equilíbrio estrutural. Usando esta nova definição, a estrutura de equilíbrio foi generalizada para uma versão rotulada *Equilíbrio Estrutural Relaxado* (DOREIAN; MRVAR, 2009). Essa nova definição, como veremos, nos dará uma nova medida de desequilíbrio e um novo problema de particionamento de grafos.

Definição 2.3.1. Um grafo de sinais $G = (V, A, s)$ é k -equilibrado relaxado se os vértices do conjunto V podem ser particionados em $S_1, S_2, \dots, S_l$, com $l \leq k$, tal que, todas as arestas dentro de um conjunto têm o mesmo sinal e todas as arestas entre dois conjuntos também têm o mesmo sinal. Isto significa que, para cada $1 \leq i \leq l$

$$A[S_i] \subseteq A^- \text{ ou } A[S_i] \subseteq A^+ \quad (2.12)$$

e para cada $1 \leq i \neq j \leq l$

$$A[S_i : S_j] \subseteq A^- \text{ ou } A[S_i : S_j] \subseteq A^+ \quad (2.13)$$

Definição 2.3.2. Seja $P = \{S_1, S_2, \dots, S_l\}$ uma l -partição de V . O desequilíbrio relaxado $RI(P)$ de P é definido como

$$RI(P) = \sum_{1 \leq i \leq l} \min\{\Omega^+(S_i, S_i), \Omega^-(S_i, S_i)\} + \sum_{1 \leq i \neq j \leq l} \min\{\Omega^+(S_i, S_j), \Omega^-(S_i, S_j)\}. \quad (2.14)$$

Observe que segundo as definições dadas $RI(P) \leq I(P)$, para toda partição P de V .

(Problema RCC) 1. Seja $G = (V, A, s)$ um grafo de sinais, w_a um peso positivo associado a cada aresta $a \in A$ e seja k um número inteiro satisfazendo $1 \leq k \leq n$. O problema de correlação de clusters relaxado é o problema de calcular uma l -partição ($l \leq k$), P de V tal que o desequilíbrio $RI(P)$ é minimizado.

Exemplo 2.3.1. Exemplo de CC e RCC. Na figura a seguir do lado esquerdo temos $I(P) = 4$ e ao lado direito $RI(P) = 1$.

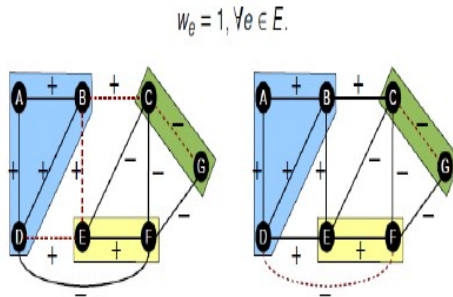


Figura 7: A diferença entre CC e RCC. Fonte (LEVORATO, 2015)

Teorema 2.3.1. *RCC é NP-difícil.*

A prova a seguir foi escrita para um grafo direcionado, pois o texto do artigo (FIGUEIREDO; MOURA, 2013) trabalha com grafo direcionado. A prova para o caso não direcionado é equivalente.

Demonstração. Para mostrar que o RCC é NP-difícil, usaremos o princípio da intratabilidade (GAREY M.R., 2003). Uma redução polinomial do CC para o RCC indica que o RCC é, no mínimo, tão difícil quanto o CC. Se o problema RCC é eficientemente resolvido por um algoritmo polinomial então o CC também será. Se o problema CC exige tempo exponencial então o RCC também exigirá.

Considere uma instância arbitrária do problema CC definida sobre um grafo de sinal direcionado $D = (V, A, s)$ e um arco do vetor de pesos w^A . Seja $D' = (V', A', s')$ um grafo sinalizado e $w^{A'}$, o vetor de pesos definido como segue

- $V' = V \cup \{n+i | i = 1, \dots, 2n\}$
- $A' = A \cup A'_1 \cup A'_2$
- $A'_1 = \{(n + (2k-1), n+2k) | k = 1, 2, \dots, n\}$ e
- $A'_2 = \{(n + (2k-1), n+i), (n+2k, n+i) | k = 1, 2, \dots, n; i = 2k+1, \dots, 2n\};$

$$s'_a = \begin{cases} s_a, a \in A, \\ +, a \in A'_1; \\ -, a \in A'_2; \end{cases}$$

$$w'_a = \begin{cases} w_a, a \in A, \\ M, a \in A'_1, \text{ onde } M = (\sum_{a \in A} w_a) + 1. \\ M, a \in A'_2; \end{cases}$$

Exemplo 2.3.2. *Exemplo para demonstração.*

- $V = \{1, 2, 3\}$
- $V' = \{1, 2, 3\} \cup \{4, 5, 6, 7, 8, 9\}$
- $A' = A \cup A'_1 \cup A'_2$
- $A'_1 = \{(4, 5), (6, 7), (8, 9)\}$ e
- $A'_2 = \{(4, 6), (4, 7), (4, 8), (4, 9), (5, 6), (5, 7), (5, 8), (5, 9), (6, 8), (6, 9), (7, 8), (7, 9)\};$

Veja a Figura 8 e o Exemplo 2.3.2 como uma ilustração. Seja $P' = \{S_1, \dots, S_n\}$ uma solução para o RCC, definido sobre o grafo D' , com $k = n$. Claramente, temos os vértices $n + (2k+1), n+2k \in S_k$, para cada $k \in \{1, \dots, n\}$. Então, o desequilíbrio dentro de cada cluster será dado pelos arcos negativos e o desequilíbrio entre diferentes clusters será dado pelos arcos positivos. Como consequência, $P' = \{S'_1, \dots, S'_n\}$ é uma solução ótima do problema CC definido sobre o grafo sinalizado D .

□

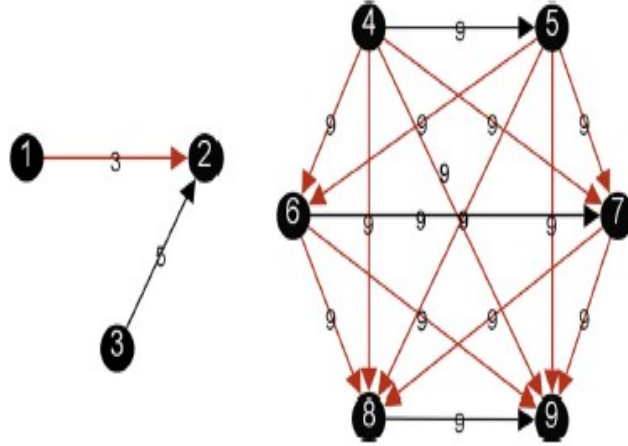


Figura 8: Exemplo da prova do teorema 2.3.1. Fonte (FIGUEIREDO; MOURA, 2013)

2.4 Modelo de PLI para o problema RCC

2.4.1 Uma formulação por representantes

A ideia principal por trás de uma formulação de representantes dada por (FIGUEIREDO; MOURA, 2013) é a representação única de um cluster pelo seu vértice com o índice mais baixo. Formulações usando representantes já foram utilizadas para resolver outros problemas, no trabalho de (CAMPELO M., 2004), primeiro a propor tal formulação e nos trabalhos de (FROTA Y., 2010) e (LEVORATO, 2015). Nesta formulação, utiliza-se variáveis de decisão binária, definidas como segue.

Para cada vértice $i \in V$ definimos:

$$\begin{cases} x_i^i = 1 \text{ se } i \text{ é um vértice representante,} \\ x_i^i = 0 \text{ caso contrário.} \end{cases}$$

Para cada par de vértices $i, j \in V$ e $i < j$ temos:

$$\begin{cases} x_j^i = 1 \text{ o vértice } j \text{ é representado pelo vértice } i \\ x_j^i = 0 \text{ caso contrário.} \end{cases}$$

Finalmente, um conjunto de variáveis binárias para representar o conjunto de arestas externas e internas.

Para cada par de vértices $i, j \in V$ e para cada aresta $(p, q) \in A$, definimos:

$$\begin{cases} y_{pq}^{ij} = 1 \text{ se } i \leq p \text{ e } j \leq q, p \text{ é representado pelo vértice } i \text{ e } q \text{ é representado por } j, \\ y_{pq}^{ij} = 0 \text{ caso contrário.} \end{cases}$$

A partir deste momento, seja S_i um conjunto de vértices representados por i . Assumiremos que $S_i = \emptyset$ quando i não é um vértice representante.

Um conjunto de variáveis contínuas é usado para descrever a função Ω^+ e Ω^- apresen-

tados na definição de desequilíbrio relaxado. Definimos $d_{ij} = \min\{\Omega^+(S_i, S_j), \Omega^-(S_i, S_j)\} \in \mathbb{R}$, para cada par $i, j \in V$. Temos a seguinte formulação:

$$\text{Minimizar } \sum_{i \leq j \in V} d_{ij} \quad (2.15)$$

sujeito a

$$\sum_{i \in V, i \leq j} x_j^i = 1, \quad \forall j \in V, \quad (2.16)$$

$$x_j^i \leq x_i^i, \quad \forall i, j \in V, \quad (2.17)$$

$$\sum_{i \in V} x_i^i = k, \quad (2.18)$$

$$y_{pq}^{ij} \in \{0, 1\} \text{ e } x_i^j \in \{0, 1\}, \quad \forall i, j \in V \text{ e } \forall p, q \in A, \quad (2.19)$$

$$y_{pq}^{ij} = x_p^i x_q^j + x_p^j x_q^i, \quad \forall i \leq j \in V, p \neq q \in A, \quad (2.20)$$

onde

$$d_{ij} = \min\left\{ \sum_{(p,q) \in A^+} w(p,q) y_{pq}^{ij}, \sum_{(p,q) \in A^-} w(p,q) y_{pq}^{ij} \right\}, \quad \forall i \leq j \in V. \quad (2.21)$$

A restrição (2.16) estabelece que o vértice j deve ser representado exatamente por um vértice representante. A restrição (2.17) proíbi o vértice j de ser representado pelo vértice i , a menos que i seja vértice representante. A restrição (2.18) define uma partição de V com no máximo k clusters. A igualdade (2.20) define um conjunto de arestas, cujas extremidades pertencem ao mesmo cluster, para $i = j$, e um conjunto de arestas ligadas por vértices de diferentes clusters, para $i \neq j$. As equações (2.15) e (2.21) definem o valor da função objetivo para o problema RCC.

A formulação do problema RCC é não linear. A restrição presente em (2.20) é quadrática, mas podemos colocá-la na forma linear, veja (WOLSEY, 1998), como segue:

$$y_{pq}^{ij} \leq x_p^i, \quad \forall i \leq j \in V, \forall (p,q) \in A., \quad (2.22)$$

$$y_{pq}^{ij} \leq x_q^j, \quad \forall i \leq j \in V, \forall (p,q) \in A., \quad (2.23)$$

$$x_q^i + x_p^j - y_{pq}^{ij} \leq 1, \quad \forall i \leq j \in V, (p,q) \in A, \quad (2.24)$$

$$0 \leq y_{pq}^{ij} \leq 1, \quad \forall i \leq j \in V, (p,q) \in A, \quad (2.25)$$

A restrição (2.24) indica que se $x_p^i = x_q^j = 1$, então $y_{pq}^{ij} = 1$. Por outro lado, as restrições (2.22), (2.23), impõem que $y_{pq}^{ij} = 0$, quando $x_p^i = 0$ ou $x_q^j = 0$.

Visto que, o problema RCC é um problema de minimização. A equação (2.21) pode ser substituída pelas seguintes disjunções:

$$\left(d_{ij} = \sum_{(p,q) \in A^+} W(p,q)y_{pq}^{ij} \right) \vee \left(d_{ij} = \sum_{(p,q) \in A^-} W(p,q)y_{pq}^{ij} \right), \forall i, j \in V. \quad (2.26)$$

A disjunção de restrições lineares pode ser equivalentemente substituída por inequações lineares escritas sobre um conjunto de variáveis binárias adicionais (WOLSEY, 1998). Note que a Equação (2.26) pode ser substituída pelas seguintes restrições:

$$d_{i,j} \leq \sum_{(p,q) \in A^+} w(p,q)y_{pq}^{ij}, \quad \forall i \leq j \in V. \quad (2.27)$$

$$d_{i,j} \leq \sum_{(p,q) \in A^-} w(p,q)y_{pq}^{ij}, \quad \forall i \leq j \in V. \quad (2.28)$$

$$d_{i,j} - \sum_{(p,q) \in A^+} w(p,q)y_{pq}^{ij} - M^+ s_{ij} \geq -M^+, \quad \forall i \leq j \in V. \quad (2.29)$$

$$d_{i,j} - \sum_{(p,q) \in A^-} w(p,q)y_{pq}^{ij} - M^-(1 - s_{ij}) \geq -M^-, \quad \forall i \leq j \in V. \quad (2.30)$$

$$s_i^j \in \{0, 1\}, \quad \forall i, j \in V, \quad (2.31)$$

$$d_i^j \geq 0, \quad \forall i \leq j \in V, \quad (2.32)$$

Onde $M^+ = \sum_{(p,q) \in A^+} w(p,q)$ e $M^- = \sum_{(p,q) \in A^-} w(p,q)$. Se $s_{ij} = 1$, a restrição (2.30) torna-se redundante e as restrições (2.27) e (2.29) forçam d_{ij} ser igual a (2.27). Por outro lado, se $s_{ij} = 0$, a restrição (2.29) torna-se redundante e as restrições (2.30) e (2.28) forçam d_{ij} ser igual a (2.28). Se ao lado direito da restrição (2.28) é menor que ao lado direito da restrição (2.27), s_{ij} é forçado a ser igual a zero. No caso oposto, isto é, se o lado direito da restrição (2.27) é menor que o lado direito da restrição (2.28), s_{ij} é forçado a ser igual a 1.

O problema RCC é equivalente ao seguinte PLI:

$$\text{Minimizar } \sum_{i \leq j \in V} d_{ij} \quad (2.33)$$

sujeito (2.16)-(2.19), (2.22)-(2.25) e (2.27)-(2.33).

2.4.2 Uma formulação usando divisão por clusters

Nesta formulação, utilizamos uma divisão por clusters para o problema RCC. Seja K o conjunto de clusters. Para cada vértice $i \in V$ e $k \in K$ com $k \leq w$, definimos:

$$\begin{cases} x_i^k = 1 \text{ se } i \text{ está no cluster } k, \\ x_i^k = 0 \text{ caso contrário.} \end{cases}$$

Um conjunto de variáveis binárias é usado para descrever o conjunto de arestas internas e externas. Para cada par de vértices $(i, j) \in A$ e para cada par de clusters $k, w \in K$, definimos:

$$\begin{cases} y_{ij}^{kw} = 1 \text{ se } i \in k \text{ e } j \in w, \\ y_{ij}^{kw} = 0 \text{ caso contrário.} \end{cases}$$

Um conjunto de variáveis contínuas é usado para descrever as funções Ω^+ e Ω^- apresentadas na definição de desequilíbrio relaxado. Definimos $d_{kw} = \min\{\Omega^+(S_i, S_j), \Omega^-(S_i, S_j)\} \in \mathbb{R}$ para cada par $k, w \in K$. Temos a seguinte formulação:

$$\text{Minimizar } \sum_{k=1}^K \sum_{w=k}^K d_{kw} \quad (2.34)$$

$$\text{Sujeito a } \sum_{k \in K} x_i^k = 1, \quad \forall i \in V, \quad (2.35)$$

$$\sum_{i \in V} x_i^k \geq 1, \quad \forall k \in K, \quad (2.36)$$

$$x_i^k \in \{0, 1\}, \quad \forall i \in V, \quad \forall k \in K \quad (2.37)$$

$$y_{ij}^{kw} = x_i^k x_j^w + x_j^k x_i^w \quad \forall (i, j) \in E, \quad \forall k \leq w \in K, \text{ onde} \quad (2.38)$$

$$d_{kw} = \min\left\{ \sum_{(i,j) \in A^+} W(i, j) y_{ij}^{kw}, \sum_{(i,j) \in A^-} W(i, j) y_{ij}^{kw} \right\} \quad \forall k \leq w \in K \quad (2.39)$$

A restrição (2.35) estabelece que o vértice i pertence exatamente a um cluster. A restrição (2.36) estabelece que nenhum cluster k está vazio, ou seja a partição terá k clusters. A restrição (2.38) define um conjunto de vértices ligados no mesmo clusters, para $i = j$ e um conjunto de vértices ligados por clusters diferentes, para $i \neq j$. As equação (2.39) e (2.34) definem o valor da função objetivo para o problema do desequilíbrio relaxado.

Como na formulação anterior, o modelo é não-linear e a linearização é análoga ao primeiro modelo.

2.4.3 Experimentos Computacionais com as formulações do RCC

Relatamos experiências computacionais realizadas com as formulações descritas neste capítulo. O objetivo destes experimentos é de estabelecer qual formulação será mais apropriada para os testes descritos no Capítulo 4. Em (FIGUEIREDO; MOURA, 2013), podemos encontrar teste para a formulação por representantes, mas não para a formulação com clusters. Faremos os testes para as instâncias presente na Tabela 1.

Note pelas tabelas a seguir (Tabelas 2, 3, 4, 5, 6 e 7) que a nova formulação para RCC por clusters apresenta resultados com tempo melhores quando $|K|$ é pequeno, onde consideramos, para estes testes, $2 \leq K \leq 7$, onde $|K|$ representa o número de clusters, I(p) o desequilíbrio, Time o tempo (em segundos) e a coluna Gap representa a diferença em porcentagem da melhor solução inteira e o limite inferior encontrado pelo método Branch-and-Bound aplicada a formulação inteira no tempo limite de $1h$.

Nome	V	A ⁺	A ⁻	soma(pesos)
House B	17	78	83	406
House C	20	93	105	466
MonkT2	18	55	49	207
MonkT3	18	57	48	207
MonkT4	18	78	76	612
MicKinney	29	246	28	264
Newcomb	17	68	51	119

Tabela 1: Literatura Instâncias

2.4.3.1 Instâncias Monastério Sampson

As instâncias conhecidas por Monastério Sampson foram definidas em 1968, (SAMPSON, 2009). Sampson recolheu dados sobre três relações assinaladas (afeição, estima, influência) para um grupo de jovens que eram ou postulantes ou novatos em um monastério em diferentes períodos de tempo.

Modelo por cluster					Modelo por representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	32.0	0.0	13.0	7301	32.0	0.0	154.0	82
3	22.0	0.0	36.0	5647	22.0	0.63	-	73041
4	16.0	0.0	101.0	10498	16.0	0.81	-	61343
5	11.0	0.0	1447.0	95389	11.0	0.81	-	53204
6	9.0	0.77	-	376172	9.0	0.88	-	52390
7	7.0	1.0	-	205217	8.0	0.75	-	66902

Tabela 2: Instância MonkT3

A partir dos resultados obtidos na Tabela 2 note que as soluções usando a formulação do problema RCC por cluster obtêm soluções ótimas ($Gap = 0$) para os valores de $K < 6$, por outro lado, na formulação por representantes o para os valores $K > 2$, a formulação não obteve solução ótima no tempo limite.

Modelo cluster					Modelo representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	25.0	0.0	10.0	2093	25.0	0.0	85.0	332
3	21.0	0.0	68.0	7331	21.0	0.66	-	56416
4	10.0	0.0	112.0	8419	10.0	0.7	-	30520
5	6.0	0.0	1068.0	55732	6.0	0.0	2161.0	47055
6	4.0	0.0	789.0	33600	4.0	0.0	2431.0	34465
7	3.0	0.99	-	113123	3.0	0.0	412.0	17614

Tabela 3: Instância MonkT4

A partir dos resultados obtidos na Tabela 3 vemos que as soluções usando a formulação do problema RCC por cluster, tem solução ótima $Gap = 0$ para $K < 7$, por outro lado, a formulação por representantes não obteve solução ótima para $K = 3, 4$.

2.4.3.2 Instâncias House

As instancias House foram geradas em 1952 por (LEMANN T.B., 1952) em um estudo sociométrico. Estas redes sociais foram obtidas de estudantes que vivem em diferentes dormitórios. Originalmente, quatro sinais de rede foram definidos, cada um associado com um indicador (encontro, amigos, colegas de quarto e final de semana).

Modelo por cluster					Modelo por representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	103	0.0	17.0	8987	103.0	0.0	63.0	4012
3	84.0	0.0	48.0	19139	84.0	0.97	-	101695
4	73.0	0.0	1369.0	427283	84.0	1.0	-	68131
5	63.0	0.79	-	623708	79.0	1.0	-	57296
6	63.0	0.87	-	425447	69.0	1.0	-	64304
7	47.0	0.80	-	442713	49.0	1.0	-	57662

Tabela 4: Instância House B

A partir dos resultados obtidos na Tabela 4, tem-se que as soluções usando a formulação do problema RCC por cluster tem solução ótima para $K < 5$. Por outro lado, na formulação por representantes temos, para $K > 2$, a formulação não obteve solução ótima no tempo limite.

Modelo por cluster					Modelo por representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	64.0	0.0	65.0	11373	64.0	0.0	371.0	5838
3	56.0	0.0	78.0	9396	60.0	1.0	-	3006
4	43.0	0.0	459.0	52386	48.0	1.0	-	13615
5	40.0	0.7	-	246926	58.0	1.0	-	11050
6	36.0	0.97	-	103631	38.0	1.0	-	9357

Tabela 5: Instância House C

A partir dos resultados obtidos na Tabela 5, vemos que as soluções usando a formulação do problema RCC por cluster obteve solução ótima para $K < 5$. Por outro lado, na formulação por representantes o para $K > 2$, não obteve solução no tempo limite.

2.4.3.3 Instância Mickney

Esta instância foi introduzida na literatura de métodos computacionais para o equilíbrio estrutural por (BRUSCO et al., 2011) que pretendiam testar sua metodologia em uma rede social grande e densa. Este digrafo assinado foi definido com base em dados coletados por (MCKINNEY, 1948) durante um estudo sobre as relações entre crianças em uma mesma sala de aula. Em um teste sociométrico, os alunos foram submetidos à escolha entre "disposto a trabalhar com outras crianças", classificado como +1, "não estar disposto a trabalhar", classificado como -1 e "indiferente", classificado como 0, como definição de seu comportamento em um grupo de discussão com os outros membros da classe.

Modelo por cluster					Modelo por representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	8.0	0.0	5.0	453	8.0	0.0	236.0	5677
3	2.0	0.0	19.0	3345	2.0	1.0	-	8792
4	0.0	0.0	8.0	525	0.0	0.0	600.0	2145
5	0.0	0.0	9.0	479	0.0	1.0	-	528
6	0.0	0.0	14.0	325	0.0	0.0	185.0	632
7	0.0	0.0	11.0	278	0.0	0.0	186.0	575

Tabela 6: Instância Mickney

A partir dos resultados obtidos na Tabela 6, vemos que as soluções usando a formulação do problema RCC por cluster têm solução ótima para todos K s. Por outro lado, na formulação por representantes, para $K = 5$, não obteve solução no tempo limite.

2.4.3.4 Instância Newcomb

A instância New Comb é uma rede social muito conhecida e assinada que foi analisada em alguns trabalhos já citados ((DOREIAN; MRVAR, 1996; DOREIAN, 2008; BRUSCO et al., 2011). Os estudantes foram recrutados para morar em casas fora do campus (fraternidade), alugado para eles como parte do Projeto de Estudo do Grupo de Michigan, supervisionado por Theodore Newcomb, de 1953 a 1956. Todos eram estudantes de transferência, ingressantes sem qualquer conhecimento prévio uns dos outros. Um estudo sociométrico com estudantes em um dormitório em uma universidade que durou várias semanas (NEWCOMB, 1961). Estes dados foram ligeiramente alterados por (DOREIAN, 2008) e aqui consideramos esta versão modificada da instância.

Modelo por cluster					Modelo por representante			
K	I(p)	Gap	Time	Nodes	I(p)	Gap	Time	Nodes
2	21.0	0.0	18.0	6170	21.0	0.0	40.0	1187
3	19.0	0.0	36.0	11704	19.0	0.89	-	114113
4	17.0	0.0	421.0	87411	18.0	0.94	-	85384
5	15.0	0.26	-	767432	18.0	0.94	-	30281
6	14.0	0.713	-	271513	14.0	0.92	-	127220
7	13.0	0.99	-	131062	18.0	0.94	-	30311

Tabela 7: Instância Newcomb

Notamos, pela Tabela 7, que para $|K|$ pequeno, a formulação com cluster obtém solução ótima e a partir de $K \geq 4$ o Gap da solução aumenta. Já a formulação por representante obtém solução ótima somente no caso $K = 2$.

Concluimos que para todas as instâncias, o tempo destinado a determinação de soluções para problema RCC, na formulação por cluster para todos os valores de K , o tempo destinado ao mesmo objetivo para formulação por representantes para K pequeno.

3 Grafos de sinais probabilístico

A seguir, definiremos o que denomina-se por função de perda do CC. Por hora, podemos sintetizá-la como a soma entre todos os pares de sinais positivos entre os clusters e todos os pares de sinais negativos internos aos clusters.

3.1 Caso probabilístico para CC

Definição 3.1.1. Dado $\mathcal{C}$ e G , a função de perda para o CC é :

$$L(\mathcal{C}, G) = |(s(i, j) = +) \wedge (i \in C_k, j \in C_l, l \neq k)| + |(s(i, j) = -) \wedge (i, j \in C_k)| \quad (3.1)$$

Exemplo 3.1.1. Considere o grafo $G = (V, A, s)$ dado pela figura 9, onde as arestas positivas e negativas são representadas pelas cores verde e vermelha tracejada, respectivamente.

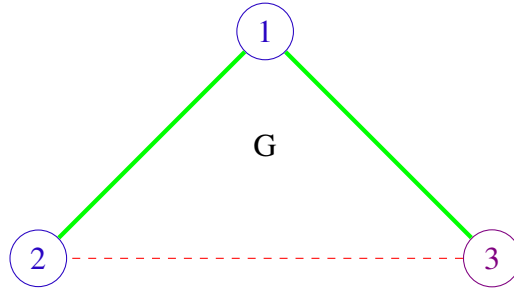


Figura 9: Grafo sinalizado G com $n = 3$.

$\mathcal{C}$ é a partição $\mathcal{C} = \{C_1, C_2\}$, com os clusters $C_1 = \{1, 2\}$ e $C_2 = \{3\}$. Então, $L(\mathcal{C}, G) = 1$

Definição 3.1.2. Denotamos por p_{ij} a probabilidade da aresta $(i, j) \in A$ ser positiva, isto é, $p(s(i, j) = +)$, onde $p_{ij}^+ = 1 - p_{ij}^-$. Assumindo que estas informações sobre as arestas são independentes, obtemos a probabilidade $P(G)$ de um grafo de sinais G ocorrer, empregando a expressão a seguir:

$$P(G) = \prod_{i, j: s(i, j) = +} p_{ij} \prod_{i, j: s(i, j) = -} (1 - p_{ij}) \quad (3.2)$$

Definição 3.1.3. Seja Θ o conjunto de todos os grafos de sinais que podem ser obtidos de V e A . A definição de Expectativa de Perda $E(\mathcal{C})$ da escolha de uma partição $\mathcal{C}$ é dada por

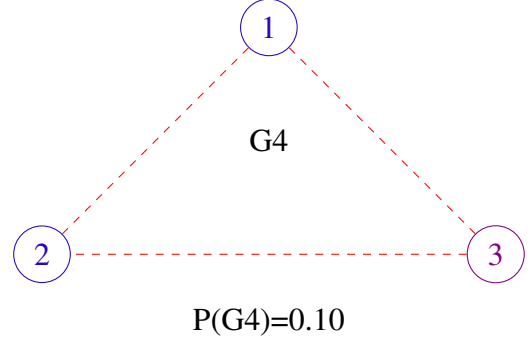
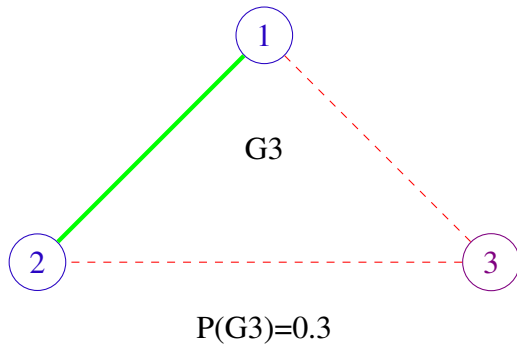
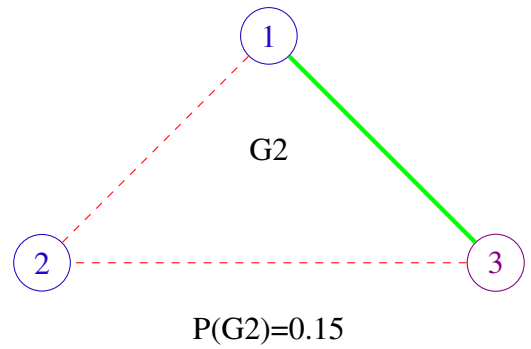
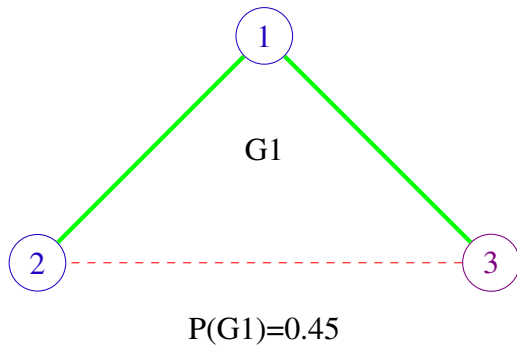
$$E(\mathcal{C}) = \sum_{G \in \Theta} L(\mathcal{C}, G) P(G) \quad (3.3)$$

onde $L(\mathcal{C}, G)$ é a perda causada pela escolha do cluster $\mathcal{C}$ no grafo G .

Exemplo 3.1.1. Do grafo de sinais G na Figura 9, consideramos a probabilidade $p_{12} = 0.75$, $p_{13} = 0.6$ e $p_{23} = 0$ (assumimos que o sinal é negativo entre os vértices 2 e 3). Então, a probabilidade do grafo G no exemplo é

$$P(G1) = p_{12} \cdot p_{13} \cdot (1 - p_{23}) = 0.75 \cdot 0.6 \cdot 1 = 0.45 \quad (3.4)$$

Neste exemplo, temos quatro grafos possíveis $\Theta = \{G1, G2, G3, G4\}$ e calculamos o valor da probabilidade para cada grafo G , dada pela equação (3.2).



Considerando agora a partição $\mathcal{C}_1 = \{C_1, C_2\}$, com os clusters $C_1 = \{1, 2\}$ e $C_2 = \{3\}$ utilizados no grafo da figura 9, temos :

$$L(\mathcal{C}_1, G1) = 1 \quad L(\mathcal{C}_1, G2) = 2 \quad L(\mathcal{C}_1, G3) = 0 \quad L(\mathcal{C}_1, G4) = 1$$

$$\text{e } E(\mathcal{C}_1) = 0.85$$

Queremos extrair a melhor partição possível para um grafo G . Ou seja, obter a partição que gera a menor perda. A comparação entre duas partições $\mathcal{C}_i$ e $\mathcal{C}_j$, pode ser obtida comparando as Esperanças de Perda causadas por estas partições.

Definição 3.1.4. Dizemos que $\mathcal{C}_j \succ \mathcal{C}_i$ se

$$E(\mathcal{C}_i) \leq E(\mathcal{C}_j)$$

Exemplo 3.1.2. Considerando agora a partição $\mathcal{C}_1 = \{C_1, C_2\}$ dada pelos clusters $C_1 = \{1, 2\}$ e $C_2 = \{3\}$ do Exemplo 3.1.1 e a partição $\mathcal{C}_2 = \{C_3, C_4\}$, onde $C_3 = \{2\}$, $C_4 = \{1, 3\}$, temos:

$$L(\mathcal{C}_2, G_1) = 1 \quad L(\mathcal{C}_2, G_2) = 0 \quad L(\mathcal{C}_2, G_3) = 2 \quad L(\mathcal{C}_2, G_4) = 1$$

$$\text{e } E(\mathcal{C}_2) = 1.15$$

Claramente, a partição $\mathcal{C}_1$ é melhor do que a partição $\mathcal{C}_2$, pois temos:

$$E(\mathcal{C}_1) \leq E(\mathcal{C}_2). \quad (3.5)$$

Nós queremos uma partição $\mathcal{C}$ que minimiza a esperança. Utilizaremos um modelo baseado no CC para determiná-la.

3.1.1 Modelo de Correlação de Clustering Probabilístico

Assumimos arestas com pesos $p \in [0, 1]$. Nesta formulação, como já vimos na Seção 2.4, uma variável de decisão binária x_{ij} é atribuída a cada par de vértices $i, j \in V, i \neq j$, e definida da seguinte maneira: $x_{ij} = 0$ se i e j estão no mesmo clusters; $x_{ij} = 1$, caso contrário. Seja $P(G)$ a probabilidade de um grafo $G \in \Theta$ ocorrer. O modelo a seguir minimiza a esperança de perda

$$\text{Minimize } \sum_{G \in \Theta} (L(\mathcal{C}, G) * P(G)) \quad (3.6)$$

$$x_{iw} + x_{wj} \geq x_{ij}, \quad \forall i, w, j \in V, \quad (3.7)$$

$$x_{ij} = x_{ji}, \quad \forall i, j \in V, \quad (3.8)$$

$$x_{ij} \in \{0, 1\}, \quad (3.9)$$

$$L(\mathcal{C}, G) = \sum_{(i,j) \in A^-} p_{i,j}(1 - x_{ij}) + \sum_{(i,j) \in A^+} p_{i,j}x_{ij}. \quad (3.10)$$

As desigualdades triangulares dadas pela equação (3.7) dizem que, se i e w estão no mesmo cluster assim como w e j , então os vértices i e j devem estar no mesmo cluster. O conjunto de restrições (3.8), escritas para todo $i, j \in V$, estabelece que as variáveis x_{ij} e x_{ji} sempre assumem o mesmo valor nesta formulação. As restrições em (3.9) impõem às variáveis o caráter binário, enquanto a função objetivo em (3.6), com a expressão para $L(C, G)$ dada em (3.10), minimiza a esperança de perda.

3.2 Caso probabilístico para o RCC

Definição 3.2.1. Seja $G = (V, A, s)$ um grafo de sinais e w_a o peso positivo associado a uma aresta $a \in A$. Então, para $1 \leq i \neq j \leq l$ tem-se

$$\overline{\Omega}^+(S_i, S_j) = \sum_{(a \in A^+) \wedge (a \in A[S_i, S_j])} w_a \quad (3.11)$$

e

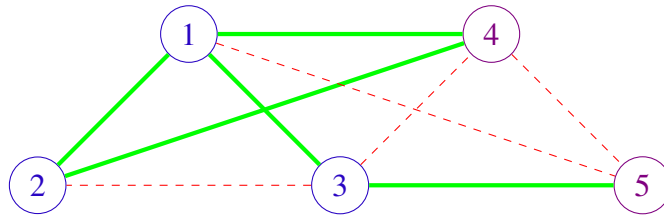
$$\overline{\Omega}^-(S_i, S_j) = \sum_{(a \in A^-) \wedge (a \in A[S_i, S_j])} w_a \quad (3.12)$$

Para o RCC, temos uma nova definição de perda. Seja $\mathcal{C} = \{C_1, C_2, \dots, C_l\}$ uma l -partição de V . A perda relaxada $LR(\mathcal{C}, G)$ é definida como

$$LR(\mathcal{C}, G) = \sum_{1 \leq i \leq l} \min\{\overline{\Omega}^+(S_i, S_i), \overline{\Omega}^-(S_i, S_i)\} + \sum_{1 \leq i \neq j \leq l} \min\{\overline{\Omega}^+(S_i, S_j), \overline{\Omega}^-(S_i, S_j)\} \quad (3.13)$$

Ou seja, o mínimo da soma do total das arestas (positivas ou negativas) predominantes dentro de cada clusters mais o mínimo da soma do total das arestas (positivas ou negativas) predominantes entre pares de clusters.

Exemplo 3.2.1. Considere um grafo $G = (V, A, s)$ dado pela figura a seguir, onde as arestas positivas e negativas são representadas pelas cores verde contínua e vermelha tracejada, respectivamente. Tomemos a partição $\mathcal{C}_1 = \{C_1, C_2\}$, com os clusters $C_1 = \{1, 2, 3\}$ e $C_2 = \{4, 5\}$.



Seguindo a definição 3.2.1, temos:

$$\begin{array}{lll} \overline{\Omega}^+(C_1, C_1) = 2 & \overline{\Omega}^-(C_2, C_2) = 1 & \overline{\Omega}^+(C_1, C_2) = 3 \\ \overline{\Omega}^-(C_1, C_1) = 1 & \overline{\Omega}^+(C_2, C_2) = 0 & \overline{\Omega}^-(C_1, C_2) = 2 \end{array}$$

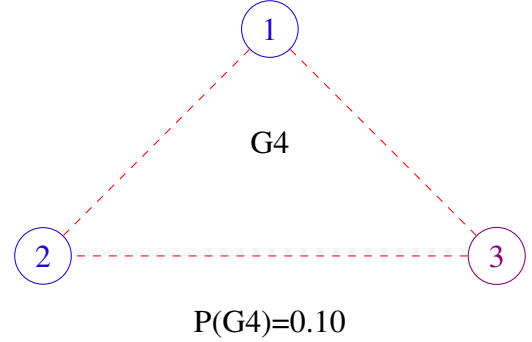
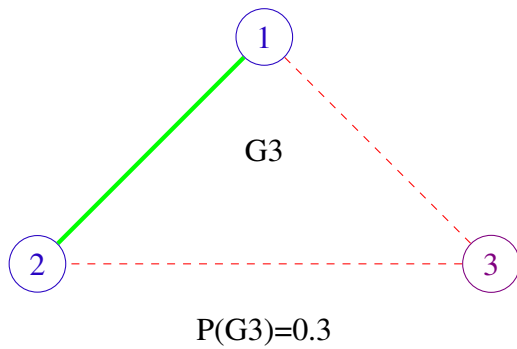
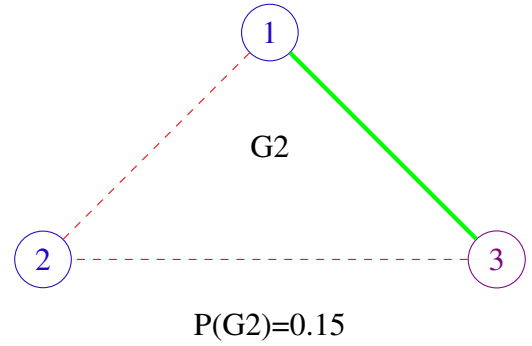
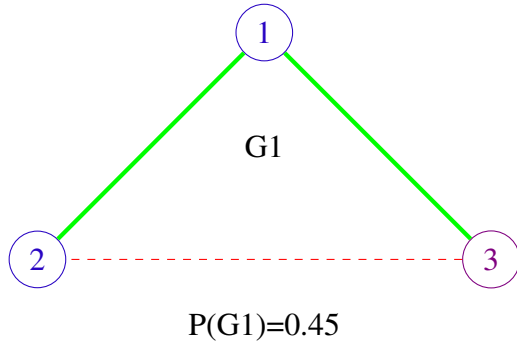
$$LR(\mathcal{C}, G) = \min\{\overline{\Omega}^+(C_1, C_1), \overline{\Omega}^-(C_1, C_1)\} + \min\{\overline{\Omega}^+(C_2, C_2), \overline{\Omega}^-(C_2, C_2)\} + \min\{\overline{\Omega}^+(C_1, C_2), \overline{\Omega}^-(C_1, C_2)\} = 1 + 0 + 2 = 3.$$

Definição 3.2.2. A expectativa de perda $E(\mathcal{C})$ de uma partição $\mathcal{C}$ é dada por

$$E(\mathcal{C}) = \sum_{G \in \Theta} LR(\mathcal{C}, G) P(G) \quad (3.14)$$

onde $LR(\mathcal{C}, G)$ é a perda relaxada causado pela partição $\mathcal{C}$ no grafo G .

Exemplo 3.2.1. Tomemos o mesmo exemplo dado pela Figura 9. Consideramos a mesma probabilidade $p_{12} = 0.75$, $p_{13} = 0.6$ e $p_{23} = 0$ (Como antes assumimos que o sinal é negativo entre os vértices 2 e 3).



Para a partição $\mathcal{C}_1 = \{C_1, C_2\}$ onde $C_1 = \{1, 2\}$, $C_2 = \{3\}$, temos:

$$L(\mathcal{C}_1, G1) = 1 \quad L(\mathcal{C}_1, G2) = 1 \quad L(\mathcal{C}_1, G3) = 0 \quad L(\mathcal{C}_1, G4) = 0$$

$$\text{e } E(\mathcal{C}_1) = 0.6$$

E para partição $\mathcal{C}_2 = \{C_3, C_4\}$ onde $C_3 = \{2\}$, $C_4 = \{1, 3\}$, temos :

$$L(\mathcal{C}_2, G1) = 1 \quad L(\mathcal{C}_2, G2) = 0 \quad L(\mathcal{C}_2, G3) = 1 \quad L(\mathcal{C}_2, G4) = 0$$

$$\text{e } E(\mathcal{C}_2) = 0.75$$

Como no caso anterior, queremos a partição que dará a menor esperança de perda relaxada para todas as possibilidades de um grafo de sinal G . Chamaremos este modelo relaxado de correlação de clusters probabilístico.

3.2.1 Modelo Relaxado de Correlação de Clustering Probabilístico

Para cada vértice $i \in V$ definimos:

$x_i^k = 1$, se i pertence ao cluster k , e $x_i^k = 0$, caso contrário.

Para cada par de aresta $(i, j) \in A$ e para cada par de clusters $k, w \in K$, teremos:

$y_{ij}^{kw} = 1$, se $i \in k$ e $j \in w$, e $y_{ij}^{kw} = 0$, caso contrário.

Continuaremos denotando por $P(G)$ a probabilidade de um grafo $G \in \Theta$ ocorrer. Por fim, um conjunto de funções contínuas é usado para definir Ω^+ e Ω^- que aparece na definição de perda relaxada (2.1.11), para um grafo e todos os grafos prováveis a partir dele. A formulação para encontrarmos a menor Expectativa de Perda relaxada é a seguinte .

$$\text{Minimizar } \sum_{G \in \Theta} \left[\left(\sum_{k=1}^K \sum_{w=k}^K LR(G)_{kw} \right) * P(G) \right] \quad (3.15)$$

$$\text{Sujeito a } \sum_{i \in V; k \in K} x_i^k = 1, \quad \forall i \in V, \quad (3.16)$$

$$\sum_{i \in V} x_i^k \geq 1, \quad \forall k \in K, \quad (3.17)$$

$$x_i^k \in \{0, 1\}, \quad \forall i \in V, \quad \forall k \in K \quad (3.18)$$

$$y_{ij}^{kw} = x_i^k x_j^w + x_j^k x_i^w \quad \forall i, j \in V, \quad \forall k \leq w \in K, \text{ onde} \quad (3.19)$$

$$LR(G)_{kw} = \min \left\{ \sum_{(i,j) \in A^+} W(i, j) y_{ij}^{kw}, \sum_{(i,j) \in A^-} W(i, j) y_{ij}^{kw} \right\} \quad \forall k \leq w \in K, \forall G \in \Theta. \quad (3.20)$$

A restrição (3.16) estabelece que o vértice i pertence exatamente a um cluster . A restrição (3.17) estabelece que nenhum cluster k está vazio, ou seja a partição terá k clusters. A restrição (3.19) define um conjunto de vértices no mesmo cluster ligados por uma aresta, para $k = w$, e define um conjunto de vértices em cluster diferentes ligados por uma aresta, para $k \neq w$. As equações (3.20) e (3.15) definem o valor da função objetivo para o problema da Esperança de Perda Relaxado.

Observe que no modelo da Seção 2.4.2 cada par de clusters k, w dava a origem a uma variável d_{kw} e consequentemente a um conjunto de variáveis binárias s_{ij} . Agora cada $LR(G)_{kw}$ da origem a uma variável d_{kw}^g e um conjunto de variáveis binárias s_{ij}^g . Isso mostrar que teremos uma dificuldade para resolver este modelo.

4 Análise de Corrupção

Nesta seção mostraremos como utilizando o CC e o RCC e suas versões probabilísticas, podemos agrupar redes de corrupção e compreender melhor a estrutura da rede de licitações em um contexto de *state capture*.

Na literatura podemos encontrar trabalhos que aplicam grafos em redes de licitações em (GOUNEV; BEZLOV, 2010) onde a pesquisa centra-se em como o crime organizado e de colarinho branco usam a corrupção para se dirigir às instituições públicas (políticos, policiais, judiciários e alfandegários), bem como ele é usado para o funcionamento de mercados-chave criminosos (drogas, roubo e extorsão). O estudo também examina como funcionários de empresas do setor privado são corrompidos por criminosos organizados. O método de agrupamento utilizado baseia-se em Redes Neurais.

4.1 O Índice de Risco de Corrupção (IRC) em *state capture*

Em uma economia capturada, os funcionários públicos vendem ou concedem, em benefício próprio bens públicos e diversas vantagens a empresas individuais. Segundo Ramos,

"A medição do comportamento corrupto é uma tarefa bastante árdua, pois a natureza intrínseca da corrupção não permite que se disponha de dados concretos sobre quão grande é sua incidência. Subornos, extorsões, desvio de verbas públicas, superfaturamento de obras, entre outras práticas ilícitas, não deixam gravados o montante desviado nem a forma como se dão tais práticas. O trabalho de oferecer dados sobre a corrupção impõe, portanto, diversas limitações. Assim a melhor maneira de se ter uma ideia sobre a incidência de práticas corruptas em uma sociedade é, portanto, trabalhar com uma percepção de sua presença. ((RAMOS, 2013),pág 22)"

Em 2014 o Centro de Investigação em Corrupção de Budapeste estabeleceu uma nova medida robusta de percepção dos riscos de corrupção de transações em contratos públicos que incidem sobre relações binárias entre emitentes e fornecedores para caso do *state capture*. Tóth e Fazekas apresentaram em seu trabalho (FAZEKAS; TÓTH; KING, 2013) o Índice de Risco de Corrupção (IRC) que mede a probabilidade de que o princípio de acesso aberto seja violado no processo de adjudicação ¹ e a execução de contratos públicos, a fim de medir a grande corrupção institucionalizada nas despesas públicas.

A probabilidade de ocorrer corrupção (ou IRC) é representada por um peso pertencente ao intervalo $[0, 1]$ ($0 \leq IRC \leq 1$), onde:

- 0 = risco de corrupção mínima ;
- 1 = risco de corrupção máxima ;

¹ é a concessão ou atribuição de direito da realização de serviços ao vencedor de uma concorrência pública, mediante licitação.

O processo de adjudicação de contratos corruptos é modelado utilizando regressão múltipla (ver (FAZEKAS; TÓTH; KING, 2013)), ligando prováveis insumos de corrupção através de informações qualitativas disponíveis nos avisos, tais como:

- processo de seleção (aberto, negociado, leilão eletrônico, preço apenas),
- número de concorrentes, alocação, publicidade, tipo de compradores (cidade, universidade, ministérios, região),
- identidade do montante disponibilizado pelo fornecedor.

Este índice mede a prevalência das irregularidades (favoritismo) nos contratos públicos, que estão associadas a uma falta de concorrência e a uma elevada quota de mercado. Em outras palavras, o índice avalia o risco de que um determinado contrato possa ser adjudicado particularmente, ou seja, não de forma justa e transparente. Neste contexto a corrupção ocorre quando uma determinada empresa já está pré-selecionada pelo contrato não visível aos outros concorrentes.

Através dos grafos de sinais discutimos as configurações exatas de rede que estamos analisando. Os grafos de sinais representam as relações entre organizações públicas e empresas privadas onde cada vínculo ou aresta representa uma relação contratual entre os dois. O sinal das arestas representam a ocorrência ou não de favoritismo no processo competitivo.

4.2 Configurações das redes de corrupção em grafos de sinais

Estas configurações foram desenhadas por (FAZEKAS; TÓTH, 2014b). A seguir, I1 -I4 representam 4 emissores diferentes (organizações públicas), S1 -S4 representam 4 diferentes fornecedores (organizações privadas), as setas verdes entre eles denotam contratos de baixo risco de corrupção (processo competitivo), e as setas vermelhas indicam contratos de alto risco de corrupção (processos de favoritismo). $C1 = \{I1, I2, S1, S2\}$ e $C2 = \{I3, I4, S3, S4\}$ são dois clusters de organizações contratantes.

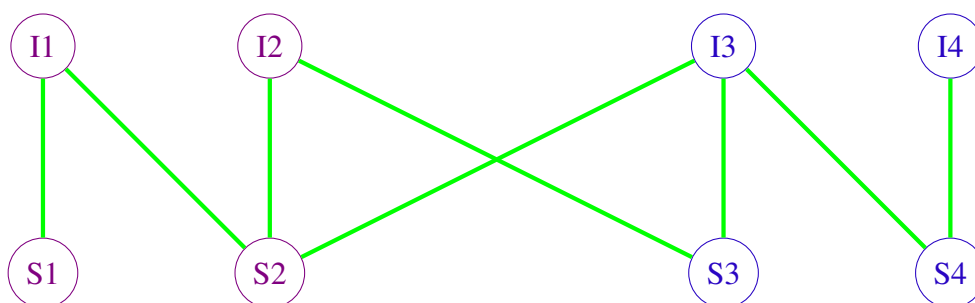
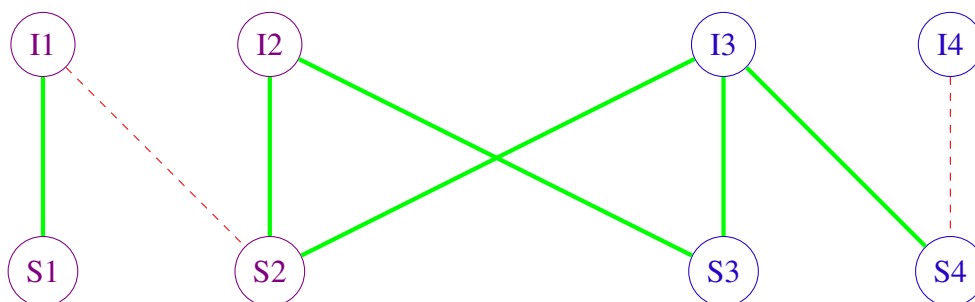


Figura 10: Estado livre de corrupção

Na Figura 10, a rede de licitações é dada por dois clusters contratuais, cada um deles com arestas de baixo índice de corrupção. Portanto, segundo (FAZEKAS; TÓTH, 2014b), a captura de estado não ocorre.

Figura 11: Corrupção sem *state capture*

Na Figura 11 temos que a corrupção exibe um padrão aleatório, sem organização em clusters de corporações corruptas. Como cada grupo de organizações contratantes tem ligações de alta e baixa corrupção, nenhum estado de captura ocorre. Segundo (FAZEKAS; TÓTH, 2014b) esse padrão aponta para fragilidades ocasionais da estrutura da rede de licitações.

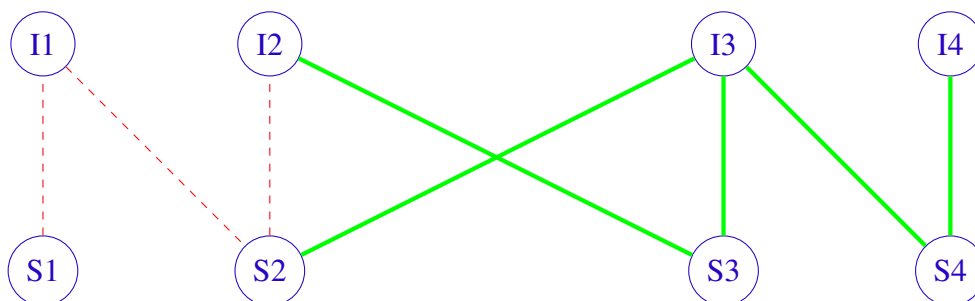


Figura 12: Estado parcialmente apropriado

A Figura 12 mostra que a corrupção é organizada no cluster (C1), que apenas abriga contratos com índices elevados de corrupção, enquanto não há corrupção no cluster (C2), que abriga somente contratos com baixo nível de corrupção. Segundo (FAZEKAS; TÓTH, 2014b), nesta situação, as organizações públicas I1 e I2 perderam toda a sua autonomia institucional no desembolso de fundos públicos através de contratação, ou seja, houve *state capture*, enquanto I3 e I4 têm conseguido manter a sua autonomia de contratação. Em uma análise ampla da rede de licitações, onde algumas organizações públicas são capturadas enquanto outras não, este estado pode ser parcialmente adequado.

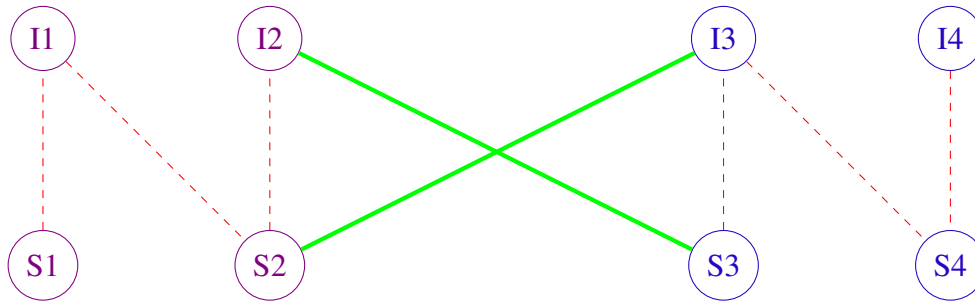


Figura 13: Estado completamente capturado

Na Figura 13 vemos que, cada contrato é de alto risco de corrupção, tornando o estado totalmente capturado, estes clusters representam uma captura global (FAZEKAS; TÓTH, 2014b). Como existem duas redes de licitações contratuais, a configuração da rede sugere uma estrutura oligárquica através da qual diferentes grupos captor alvo estão em conjuntos diferentes de corporações.

4.3 Experimentos Computacionais em Redes de Corrupção

Nesta seção, faremos o estudo de agrupamento em redes de licitações. As redes de licitações são representadas por grafos de sinais. Para dados de estudo, utilizamos redes de licitações run1 e run2 (formuladas por um especialista da Universidade de Avignon, para um possível caso de corrupção). Como amostra piloto simulamos 5 grafos aleatórios. Todas as instâncias possuem 12 vértices e 14 arestas. Onde os vértices representam entidades publicas e fornecedores, e as arestas representam as relações entre os dois.

Cada aresta recebe um peso entre 0.5 e 1, refletindo a relação entre as organizações publicas e privadas. Isto significa que todos os pesos são escalados, comparados aos resultados da corrupção. Esses pesos servem para indicar o nível de correlação entre as arestas.

- Probabilidade com valor no intervalo $[0.9, 1]$, positiva ou negativa, indica correlação muito forte.
- Probabilidade com valor no intervalo $]0.7, 0.9[$, positiva ou negativa, indica correlação forte.
- Probabilidade com valor no intervalo $[0.5, 0.7[$, positiva ou negativa, indica correlação moderada.

Exemplo 4.3.1. *Particularmente, se*

- uma aresta verde recebe peso 0.8 então os seu $IRC = 0.5 * 0.8 = 0.4$;
- uma aresta vermelha recebe peso 0.6 então seu $IRC = 0.5 * 0.6 + 0.5 = 0.8$.

Algoritmo

O algoritmo abaixo descreve como fazemos a análise a partir de pesos probabilísticos.

- **Passo 1.** Tomamos um grafo de sinais com as possíveis relações entre iminentes e fornecedores e um conjunto de pesos probabilísticos P_i para as arestas deste grafo.

P1	0.6	0.8	0.7	0.8	0.9	0.8	0.7	0.7	1.0	0.6	0.6	0.8	0.7	0.7
P2	0.9	0.6	0.5	0.7	0.6	0.8	0.8	0.7	0.5	0.7	0.6	0.9	0.6	0.7
P3	0.5	0.7	0.7	0.7	0.7	0.7	0.9	0.9	0.6	0.9	0.9	0.7	0.9	0.7
P4	0.9	0.9	0.9	0.5	0.6	0.5	0.8	0.6	0.7	0.8	0.8	0.9	0.9	0.6
P5	1.0	0.6	0.8	0.5	0.9	0.9	1.0	0.8	1.0	0.8	0.9	0.8	0.8	0.9
P6	0.6	0.7	0.8	0.7	1.0	0.8	0.9	0.6	0.8	0.9	0.8	0.7	0.9	0.9
P7	0.9	0.6	0.7	0.6	0.7	1.0	0.9	1.0	1.0	0.6	0.7	0.7	0.5	0.6
P8	0.7	0.8	0.5	0.6	0.8	0.8	0.8	0.9	0.7	0.8	0.9	0.7	0.6	1.0
P9	1.0	0.6	0.7	1.0	0.7	0.6	0.8	0.8	0.6	0.7	0.9	0.5	0.6	0.8
P10	0.9	1.0	0.6	0.7	0.7	0.9	0.6	0.6	0.9	0.8	0.7	0.6	0.8	0.7

Tabela 8: Tabelas de probabilidade

- **Passo 2.** Usamos a formulação do CC (ou RCC) para encontrar a partição que têm o menor desequilíbrio.
- **Passo 3.** Geramos todos os grafos possíveis, a partir de G . (Gerar todos grafos possíveis equivale a gerar 2^m grafos, onde m é o número de aresta). Note que, para instâncias maiores podemos utilizar os grafos mais prováveis.
- **Passo 4.** A partir da partição dado pelo passo 2. Usaremos a formulação do CC probabilístico (ou RCC probabilístico) para calcular a esperança de perda mínima da partição obtida na solução do passo 2.
- **Passo 5** Usamos a formulação CC probabilístico e o RCC probabilístico para cacular a partição que nos dará a esperança de perda mínima.
- **Passo 6.** A partir da partição $\mathcal{C}$ dada pelo passo 5, calculamos o desequilíbrio desta partição segundo $I(P)$ e $RI(P)$.

Assim, teremos duas partições uma dada pela formulação do CC (ou RCC) e outra dada pelo CCP (ou RCCP).

4.3.1 Estudo das Intâncias

Como o que queremos é encontrar os clusters com corrupção, faremos uma mudança na função de perda do CC. A soma de todos os pares de sinais negativos entre os clusters e de todos os pares de sinais positivos no mesmo cluster representam a perda.

Definição 4.3.1. A nova função de perda para o CC é:

$$L(\mathcal{C}, G) = |(x_{i,j}^G = -) \wedge (i \in C_k, j \in C_l, l \neq k)| + |(x_{i,j}^G = +) \wedge (i, j \in C_k)|. \quad (4.1)$$

A tabela a seguir apresenta os principais resultados do experimento em redes de licitações. A coluna Matrix indica as instâncias usadas. A segunda coluna mostra os conjuntos de pesos probabilísticos presente na Tabela 8. A terceira coluna o tempo de geração dos grafos possíveis a partir das instâncias. Em seguida a instâncias são divididas em duas tabelas que representa as formulações: CC, CC probabilístico, RCC e RCC probabilístico. Onde k é o número de cluster da solução. Time é tempo destinado à obtenção de solução para cada formulação (em segundos). As soluções têm como tempo limite de uma hora. Lembrando que a solução dos modelos é uma partição. Temos para cada bloco:

1. I1(sol) desequilíbrio dada pela formulação do CC e E1(sol) a esperança de perda obtida através da partição dada por esta solução.
2. E2(sol) a esperança de perda dada pelo CC probabilístico e I1(sol) a solução do CC da partição do CC probabilístico.
3. RI3(sol) perda dada pela formulação do RCC e RE1(sol) a esperança de perda obtida através da partição dada pelo RCC
4. RE4(sol) a esperança de perda dada pelo RCC probabilístico e RI4(sol) a solução do CC da partição do CC probabilístico.

Observamos pelos resultados com as instâncias que há dificuldade de solução do problema probabilístico. No caso do CCP, a dificuldade do problema probabilístico é o aumento de número de termos na função objetivo, mas mantem-se o conjunto de variáveis e restrições iguais. No caso do RCC probabilístico temos uma explosão no tamanho do modelo causadas pelas restrições 3.20, não tendo solução no tempo limite.

As soluções para o CC, CC probabilístico e do RCC são bem próximas e em alguns casos iguais.

4.3.2 Análise das instâncias

A seguir, I1 -I3 representam 3 emissores diferentes (organizações públicas), S11 -S33 representam 12 diferentes fornecedores (organizações privadas), as setas verdes contínuas entre eles denotam contratos de baixo risco de corrupção (processo competitivo), e as setas vermelhas tracejadas indicam contratos de alto risco de corrupção (processos de favoritismo) como nas configurações que foram desenhadas por (FAZEKAS; TÓTH, 2014b). Utilizamos as formulações (CC, CC probabilístico, RCC e RCC probabilístico) para obter uma partição com a menor perda e a menor expectativa de perda, a partir dos pesos presentes na Tabela 8. O grafo de sinais resultante mostrará a partição com a perda mínima. A partir disso, será possível identificar:

- as "organizações limpas": média de corrupção baixa;
- as organizações capturadas parcialmente: organizações ocasionalmente corruptas, com média de corrupção baixa e desempenho altamente variável na prática de corrupção;
- as organizações com captura total: média de corrupção alta, com baixa variabilidade de desempenho, indicando que trocas corruptas representam a norma na prática de contratação das corporações.

As melhores soluções foram dada pela formulação do CC. Por este fator, faremos análise em Run1 para confirmar os resultados numéricos. Uma análise a partir da teoria *state capture* usando a formulação do CC para Run1.

A instância Run1 na figura é vista como Estado parcialmente apropriado para uma visão global da rede. Observamos que os três iminentes I1,I2,I3 foram capturados nas duas instâncias. Nesta situação, as organizações públicas presentes nos clusters violeta possuem altos riscos de corrupção (Na maioria dos contratos houve favoritismo) tornando o estado totalmente Capturado.

4.3.2.1 Análise Run1

Matrix	Instance		CC					Probabilistic CC			
	Prob	Time(sec)	k	$I1(sol)$	Time(sec)	$EI1(sol)$	Time(sec)	k	$EI2(sol)$	Time(sec)	$I2(sol)$
run1	P1	13.01	2	0.6	0.01	1.62	17.0	2	1.62	12.91	0.6
	P2	43.98	3	0.5	0.01	2.03	46.0	2	1.69	43.36	1.3
	P3	49.19	2	0.5	0.01	1.43	66.0	3	1.06	58.68	0.6
	P4	75.2	2	0.5	0.01	1.41	86.0	3	1.13	75.44	0.7
	P5	19.05	2	0.5	0.01	1.36	29.0	2	0.99	29.16	1.1
	P6	59.24	3	0.6	0.02	1.24	28.0	3	1.24	19.89	0.6
	P7	6.64	2	0.6	0.01	1.47	44.0	2	1.33	16.13	1.2
	P8	35.07	2	0.6	0.01	1.6	88.0	3	1.39	86.25	0.7
	P9	25.98	3	0.6	0.01	1.58	95.0	3	1.29	97.51	1.2
	P10	36.5	2	0.7	0.01	1.98	112.0	3	1.67	107.58	1.5
run2	P1	35.39	3	0.0	0.01	1.42	32.0	3	1.42	18.2	0.0
	P2	116.57	3	0.0	0.02	1.53	61.0	3	1.34	57.19	0.6
	P3	74.79	3	0.0	0.01	0.93	51.0	3	0.93	50.28	0.0
	P4	113.05	3	0.0	0.01	0.91	84.0	3	0.91	84.94	0.0
	P5	21.95	3	0.0	0.01	0.86	32.0	3	0.49	25.29	0.6
	P6	44.98	3	0.0	0.02	1.04	78.0	3	1.04	72.3	0.0
	P7	10.29	3	0.0	0.01	1.28	8.0	3	1.14	7.08	0.6
	P8	29.05	3	0.0	0.01	1.14	28.0	3	1.14	26.42	0.0
	P9	16.97	3	0.0	0.01	1.46	26.0	3	1.18	21.79	0.6
	P10	33.35	3	0.0	0.01	1.47	39.0	3	1.24	37.22	0.6
A1	P1	13.25	3	0.0	0.01	1.42	17.0	3	1.42	12.89	0.0
	P2	55.77	3	0.0	0.01	1.53	56.0	3	1.34	47.44	0.6
	P3	64.91	3	0.0	0.01	0.93	64.0	3	0.93	57.53	0.0
	P4	74.51	3	0.0	0.01	0.91	84.0	3	0.91	74.44	0.0
	P5	18.85	3	0.0	0.01	0.86	28.0	3	0.49	28.59	0.6
	P6	60.01	3	0.0	0.02	1.04	70.0	3	1.04	66.05	0.0
	P7	6.34	3	0.0	0.01	1.28	44.0	3	1.14	16.35	0.6
	P8	34.77	3	0.0	0.01	1.14	89.0	3	1.14	85.03	0.0
	P9	24.27	3	0.0	0.01	1.46	85.0	3	1.18	70.85	0.6
	P10	32.29	3	0.0	0.01	1.47	100.0	3	1.34	95.86	1.2
A2	P1	35.7	2	0.6	0.01	1.62	49.0	3	1.54	49.33	0.6
	P2	90.82	3	0.7	0.01	1.74	81.0	2	1.69	73.51	1.3
	P3	96.65	2	0.5	0.01	1.43	37.0	3	1.34	34.85	0.9
	P4	110.32	3	0.8	0.01	1.22	56.0	3	1.22	51.91	0.8
	P5	23.0	3	0.8	0.01	1.23	18.0	2	0.99	18.06	1.1
	P6	69.44	3	0.6	0.01	1.24	82.0	3	1.24	69.23	0.6
	P7	11.81	3	0.6	0.01	1.4	13.0	2	1.33	4.66	1.2
	P8	34.85	2	0.7	0.01	1.54	29.0	3	1.47	28.19	0.8
	P9	25.3	3	0.7	0.01	1.67	37.0	3	1.67	34.17	0.7
	P10	43.66	3	0.8	0.01	1.79	57.0	3	1.56	56.17	1.4
A3	P1	28.25	3	0.7	0.01	1.93	17.0	5	1.64	12.37	0.8
	P2	62.28	3	0.5	0.01	2.03	46.0	3	1.34	42.82	0.6
	P3	74.84	3	0.6	0.01	1.06	67.0	3	1.06	58.92	0.6
	P4	95.79	8	0.5	0.01	1.41	87.0	4	0.95	77.54	0.9
	P5	22.23	3	0.6	0.01	0.49	29.0	3	0.49	29.16	0.6
	P6	61.42	3	0.7	0.02	1.15	92.0	3	1.15	64.69	0.7
	P7	9.74	5	0.6	0.01	1.14	44.0	3	1.14	16.43	0.6
	P8	44.94	3	0.5	0.01	1.64	90.0	5	1.25	85.68	0.8
	P9	28.91	3	0.6	0.01	1.58	86.0	3	1.18	68.52	0.6
	P10	61.02	3	0.6	0.01	1.24	101.0	3	1.24	96.45	0.6
A4	P1	21.8	4	0.6	0.01	1.54	17.0	4	1.54	10.91	0.6
	P2	61.26	4	0.5	0.01	2.03	45.0	3	1.69	43.52	1.4
	P3	87.59	4	0.6	0.02	1.06	45.0	4	1.06	39.96	0.6
	P4	91.76	4	0.5	0.01	1.41	63.0	4	1.13	63.89	0.7
	P5	24.73	4	0.8	0.01	1.23	20.0	3	1.09	20.44	1.5
	P6	40.49	3	0.8	0.02	1.37	65.0	3	1.37	61.16	0.8
	P7	17.78	4	0.6	0.01	1.4	4.0	4	1.4	3.12	0.6
	P8	37.42	4	0.7	0.01	1.39	20.0	4	1.39	19.33	0.7
	P9	16.53	4	0.6	0.01	1.58	19.0	3	1.38	14.93	1.2
	P10	37.15	4	0.7	0.01	1.88	32.0	3	1.56	40.12	1.4
A5	P1	30.75	3	0.7	0.01	1.62	21.0	3	1.62	13.65	0.7
	P2	69.71	3	0.6	0.01	1.34	50.0	3	1.34	45.14	0.6
	P3	112.48	2	0.7	0.01	1.62	73.0	3	1.33	69.74	0.9
	P4	109.32	4	0.5	0.01	1.41	97.0	3	0.95	112.5	0.9
	P5	11.76	4	0.5	0.02	1.36	8.0	3	0.49	4.63	0.6
	P6	47.35	3	0.7	0.01	1.15	73.0	3	1.15	72.47	0.7
	P7	9.08	3	0.5	0.01	1.78	14.0	3	1.14	7.03	0.6
	P8	42.79	3	0.6	0.01	1.24	45.0	3	1.24	36.91	0.6
	P9	22.58	3	0.6	0.01	1.18	9.0	3	1.18	6.2	0.6
	P10	29.17	2	0.7	0.01	2.15	22.0	4	1.55	20.51	1.4

Tabela 9: Resultados CC e CC probabilístico

Instance			RCC					Probabilistic RCC				
Matrix	Prob	Time(sec)	k	$RI3(sol)$	Time(sec)	$\overline{REI3(sol)}$	Time(sec)	k	$REI4(sol)$	Time(sec)	$\overline{RI4(sol)}$	Gap
run1	P1	13.01	2	0.0	0.01	1.34	71.0	2	2.63	-	3.5	1
	P2	43.98	2	0.0	0.05	1.46	1092.0	2	2.61	-	3.3	1
	P3	49.19	2	0.0	0.11	0.83	352.0	2	2.83	-	3.7	1
	P4	75.2	2	0.0	0.02	0.91	728.0	2	2.53	-	3.4	1
	P5	19.05	2	0.0	0.02	0.86	31.0	2	0.86	-	0.0	0.9
	P6	59.24	2	0.0	0.04	1.25	485.0	2	4.06	-	4.2	1
	P7	6.64	2	0.0	0.02	1.12	57.0	2	1.04	-	0.0	0.9
	P8	35.07	2	0.0	0.05	1.11	320.0	2	2.51	-	3.5	1
	P9	25.98	2	0.0	0.03	1.44	161.0	2	1.44	-	0.0	1
	P10	36.5	2	0.0	0.03	1.44	437.0	2	2.85	-	3.8	1
run2	P1	35.39	2	0.0	0.15	1.38	121.0	2	2.43	-	2.8	1.0
	P2	116.57	2	0.0	0.04	1.33	87.0	2	2.34	-	2.1	1
	P3	74.79	2	0.0	0.02	0.87	109.0	2	1.67	-	2.0	1
	P4	113.05	2	0.0	0.01	0.86	114.0	2	2.83	-	4.3	1
	P5	21.95	2	0.0	0.01	0.86	11.0	2	0.39	-	0.6	0.6
	P6	44.98	2	0.0	0.03	1.21	51.0	2	2.27	-	1.9	1
	P7	10.29	2	0.0	0.02	1.28	25.0	2	1.14	-	0.6	0.9
	P8	29.05	2	0.0	0.03	1.11	77.0	2	2.71	-	2.9	1
	P9	16.97	2	0.0	0.03	1.26	24.0	2	1.25	-	1.2	0.9
	P10	33.35	2	0.0	0.02	1.35	53.0	2	2.14	-	1.6	1
A1	P1	13.25	2	0.6	0.04	1.46	54.0	2	2.26	-	2.6	1
	P2	55.77	2	0.7	0.08	1.59	242.0	2	2.97	-	3.7	1
	P3	64.91	2	0.7	0.23	1.13	361.0	2	2.77	-	4.3	1
	P4	74.51	2	0.5	0.04	1.27	93.0	2	2.75	-	4.8	1
	P5	18.85	2	0.8	0.06	1.23	10.0	2	0.83	-	1.4	0.6
	P6	60.01	2	0.7	0.04	1.2	62.0	2	1.36	-	1.4	1
	P7	6.34	2	0.6	0.02	1.35	10.0	2	1.19	-	1.2	0.6
	P8	34.77	2	0.8	0.04	1.5	49.0	2	1.62	-	0.8	1
	P9	24.27	2	0.6	0.03	1.51	23.0	2	1.24	-	1.3	0.7
	P10	32.29	2	0.8	0.06	1.7	64.0	2	2.6	-	4.0	1
A2	P1	35.7	2	0.6	0.09	1.44	124.0	2	2.73	-	2.8	1.0
	P2	90.82	2	0.9	0.04	1.77	158.0	2	2.06	-	3.1	1
	P3	96.65	2	0.5	0.04	1.29	158.0	2	1.92	-	3.1	1
	P4	110.32	2	0.9	0.06	1.58	222.0	2	2.63	-	4.4	1
	P5	23.0	2	1.0	0.04	1.85	9.0	2	0.99	-	1.1	0.9
	P6	69.44	2	0.6	0.12	1.15	114.0	2	2.07	-	2.0	1
	P7	11.81	2	0.9	0.07	1.93	16.0	2	1.29	-	1.2	0.9
	P8	34.85	2	0.7	0.08	1.29	79.0	2	2.43	-	1.6	1
	P9	25.3	2	1.0	0.12	2.25	81.0	2	1.97	-	2.6	1
	P10	43.66	2	0.9	0.08	2.07	142.0	2	2.45	-	2.1	1
A3	P1	28.25	2	0.7	0.05	1.85	48.0	2	2.37	-	3.0	1
	P2	62.28	2	0.5	0.11	1.84	834.0	2	2.3	-	1.9	1
	P3	74.84	2	0.6	0.04	1.03	165.0	2	2.31	-	2.7	1
	P4	95.79	2	0.7	0.04	1.13	298.0	2	1.88	-	2.7	1
	P5	22.23	2	0.6	0.08	0.49	18.0	2	0.49	-	0.6	0.8
	P6	61.42	2	0.7	0.11	1.13	297.0	2	2.72	-	3.0	1
	P7	9.74	2	0.6	0.17	1.13	16.0	2	1.12	-	0.6	0.6
	P8	44.94	2	0.7	0.04	1.36	57.0	2	1.53	-	1.9	1
	P9	28.91	2	0.6	0.07	1.56	24.0	2	1.48	-	0.6	0.9
	P10	61.02	2	0.7	0.04	1.85	81.0	2	2.17	-	3.6	1
A4	P1	21.8	2	0.9	0.05	1.86	85.0	2	2.32	-	1.3	1
	P2	61.26	2	0.5	0.21	1.91	365.0	2	2.5	-	3.4	1
	P3	87.59	2	0.6	0.07	0.99	245.0	2	3.02	-	3.9	1
	P4	91.76	2	0.6	0.03	1.21	194.0	2	2.43	-	3.0	1
	P5	24.73	2	0.9	0.04	1.52	14.0	2	1.09	-	2.5	0.9
	P6	40.49	2	0.8	0.05	1.41	68.0	2	2.78	-	2.6	1
	P7	17.78	2	0.7	0.13	1.5	26.0	2	1.45	-	1.3	0.9
	P8	37.42	2	0.7	0.04	1.36	57.0	2	1.53	-	1.9	1
	P9	16.53	2	0.6	0.07	1.56	24.0	2	1.48	-	0.6	0.9
	P10	37.15	2	0.7	0.04	1.85	81.0	2	2.17	-	3.6	1
A5	P1	30.75	2	0.7	0.17	1.86	187.0	2	2.53	-	1.4	1
	P2	69.71	2	0.5	0.07	1.91	295.0	2	3.16	-	4.0	1
	P3	112.48	2	0.6	0.05	1.03	221.0	2	2.73	-	3.4	1
	P4	109.32	2	0.6	0.04	1.46	389.0	2	2.74	-	4.2	1
	P5	11.76	2	0.6	0.19	0.47	24.0	2	0.47	-	0.6	0.5
	P6	47.35	2	0.7	0.03	1.25	75.0	2	2.28	-	3.0	1
	P7	9.08	2	0.5	0.03	1.73	11.0	2	1.13	-	0.6	0.5
	P8	42.79	2	0.6	0.03	1.16	68.0	2	1.94	-	1.8	1
	P9	22.58	2	0.6	0.04	1.06	24.0	2	1.23	-	1.3	0.9
	P10	29.17	2	0.7	0.04	2.05	67.0	2	2.11	-	2.0	1

Tabela 10: Resultados RCC e RCC probabilístico

Matrix run1	Instance		CC					Probabilistic CC			
	Prob	Time(sec)	k	$I1(sol)$	Time(sec)	$\overline{EI1(sol)}$	Time(sec)	k	$EI2(sol)$	Time(sec)	$\overline{I2(sol)}$
	P1	13.01	2	0.6	0.01	1.62	17.0	2	1.62	12.91	0.6
	P2	43.98	3	0.5	0.01		46.0	2	1.69	43.36	1.3
	P3	49.19	2	0.5	0.01	1.43	66.0	3	1.06	58.68	0.6
	P4	75.2	2	0.5	0.01	1.41	86.0	3	1.13	75.44	0.7
	P5	19.05	2	0.5	0.01	1.36	29.0	2	0.99	29.16	1.1
	P6	59.24	3	0.6	0.02	1.24	28.0	3	1.24	19.89	0.6
	P7	6.64	2	0.6	0.01	1.47	44.0	2	1.33	16.13	1.2
	P8	35.07	2	0.6	0.01	1.6	88.0	3	1.39	86.25	0.7
	P9	25.98	3	0.6	0.01	1.58	95.0	3	1.29	97.51	1.2
	P10	36.5	2	0.7	0.01	1.98	112.0	3	1.67	107.58	1.5

Tabela 11: Resultados CC e CC probabilístico Run1

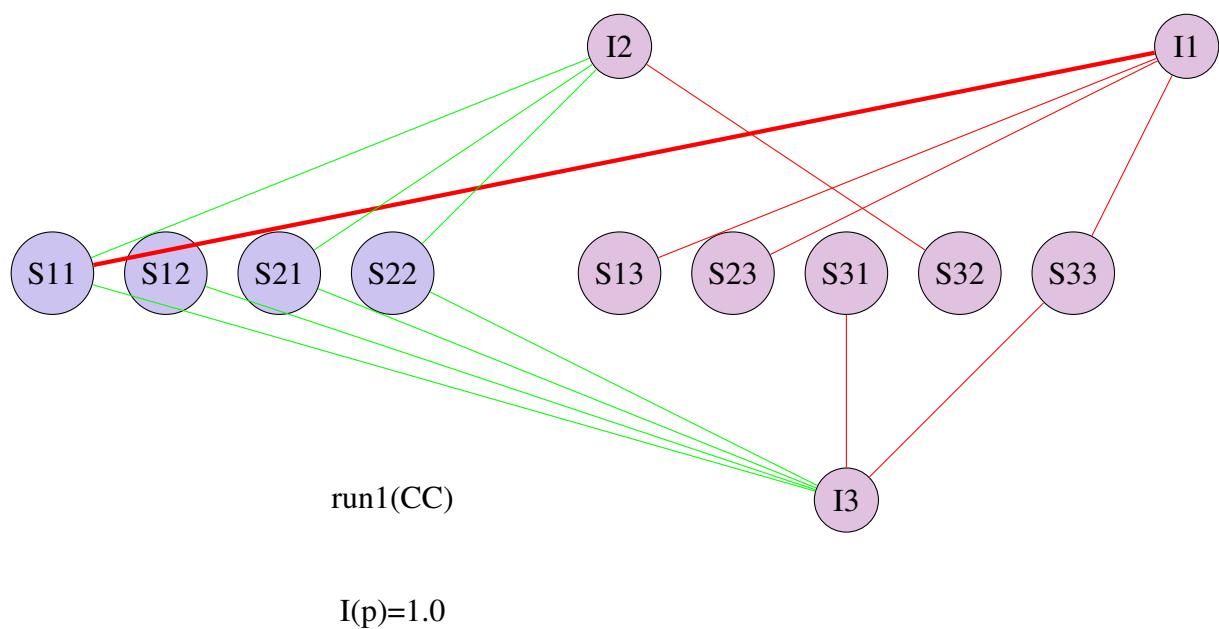
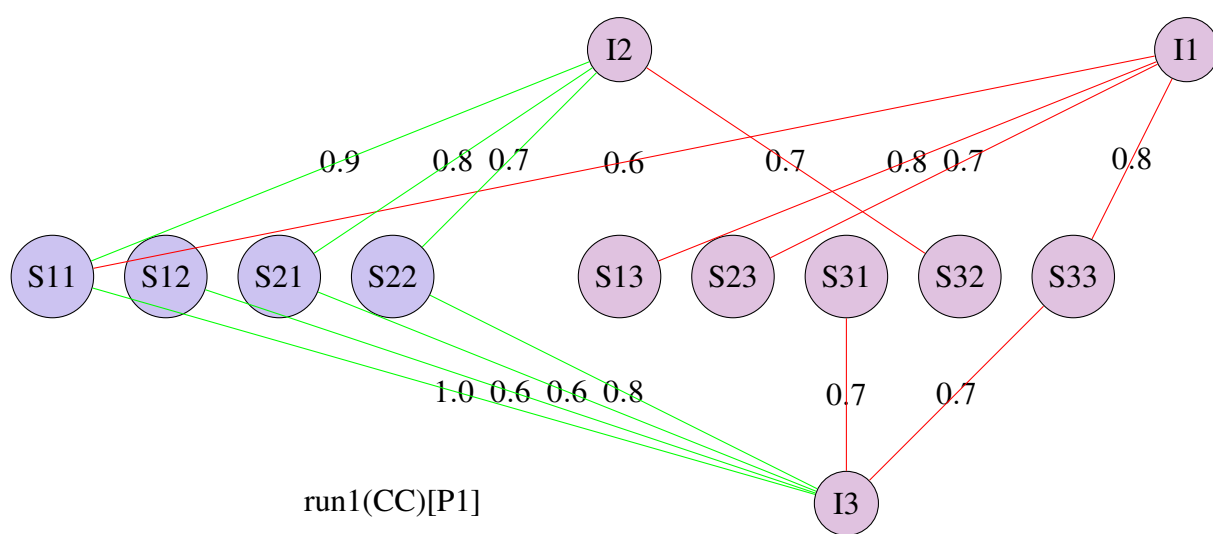
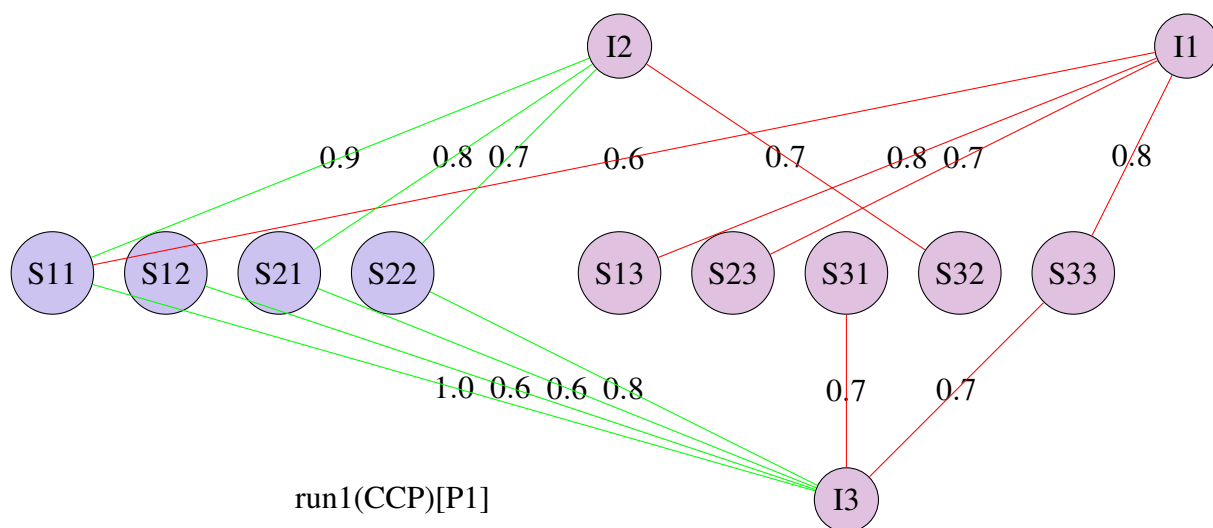


Figura 14: Run1

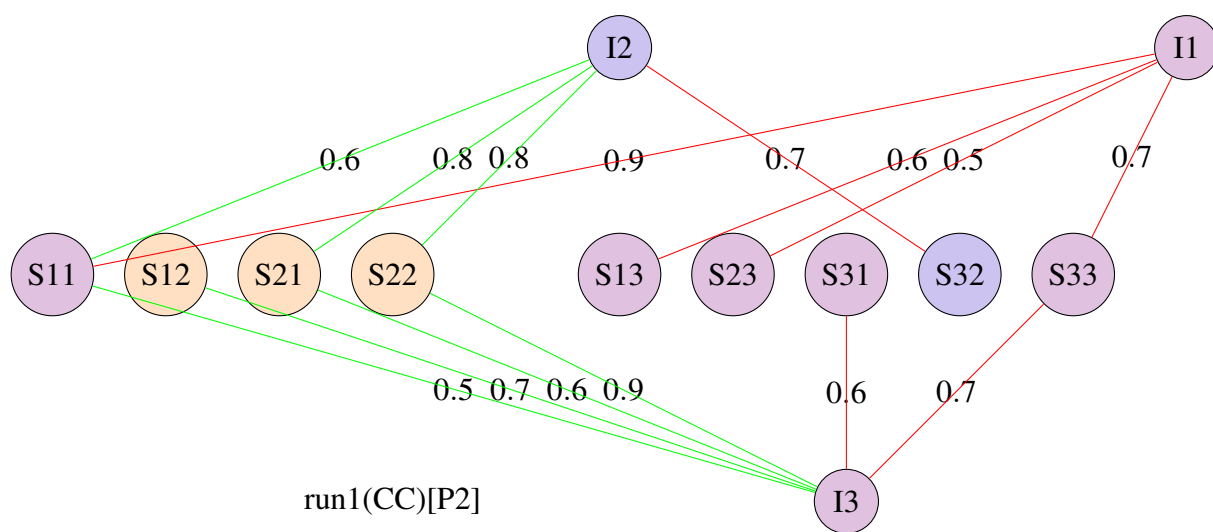


$$I1(p)=0.6 \text{ e } EI1(sol)=1.62$$

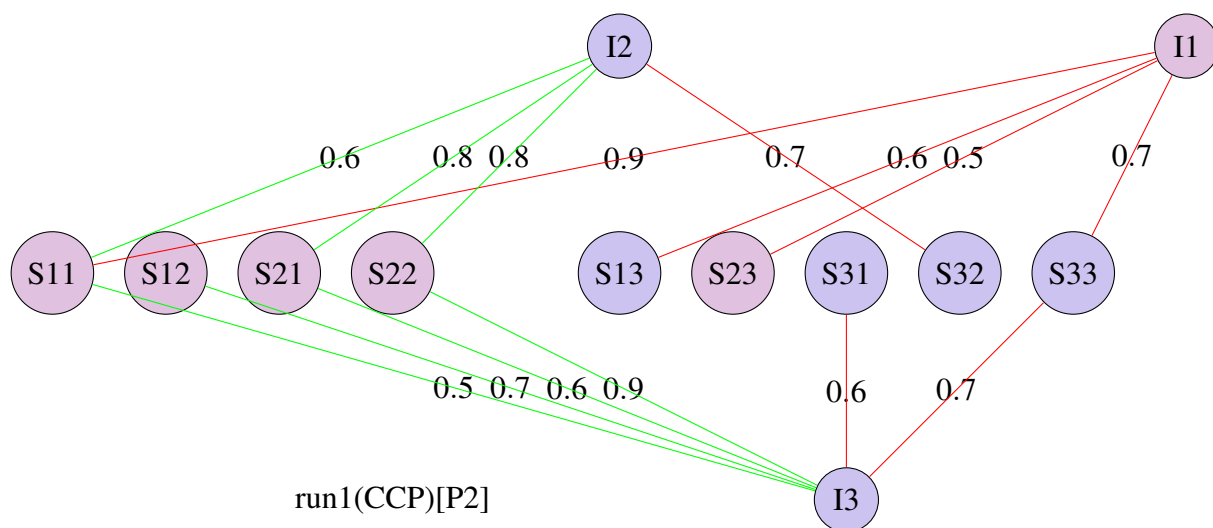


$$I1(p)=0.6 \text{ e } EI1(sol)=1.62$$

Figura 15: Run1[P1]

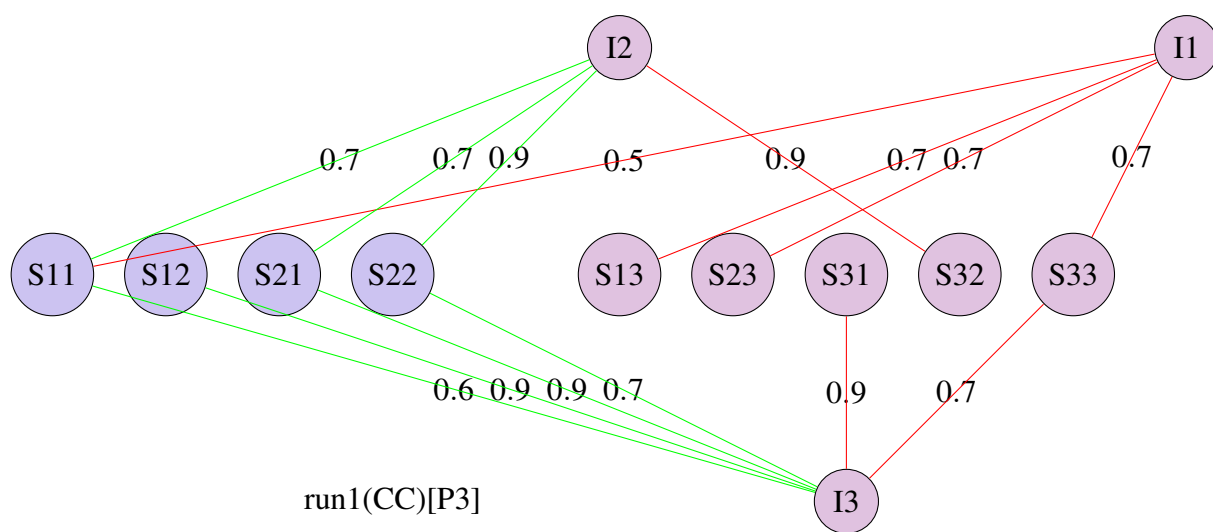


$I1(p)=0.5$ e $EI1(sol)=2.03$

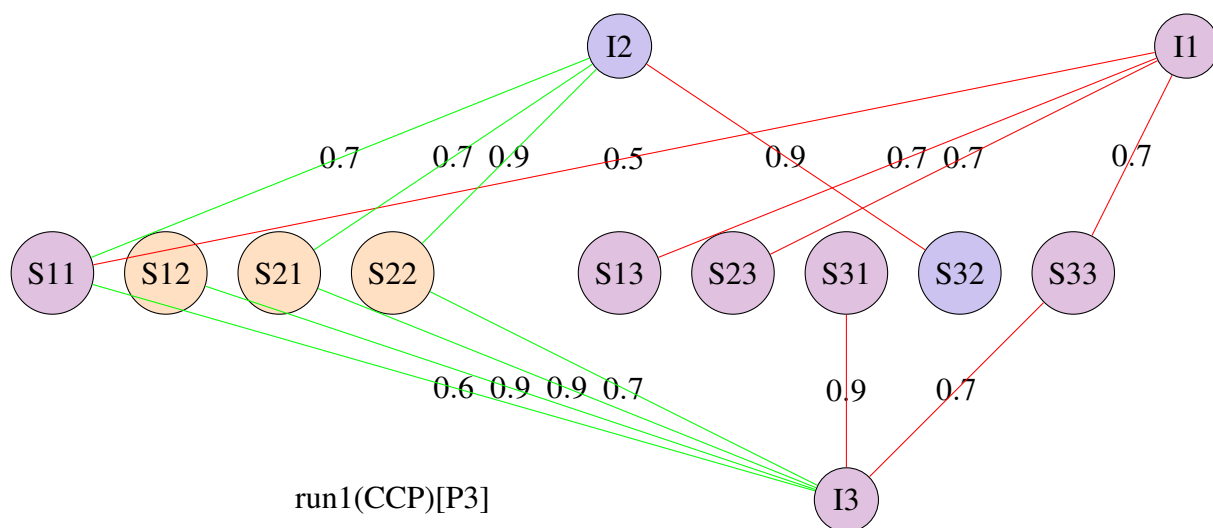


$I2(p)=1.3$ e $EI2(sol)=1.69$

Figura 16: Run1[P2]

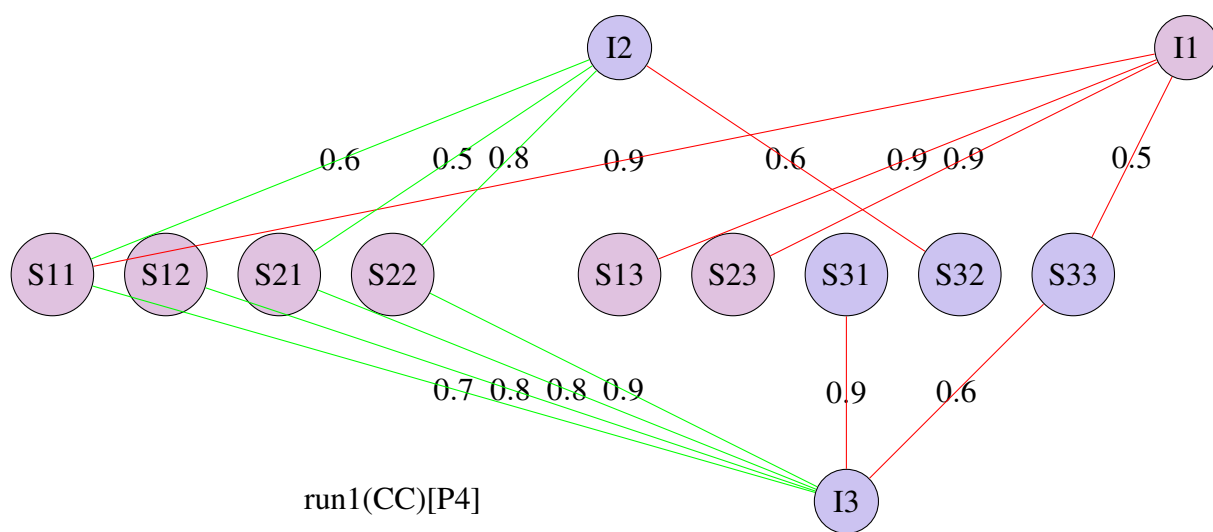


$I1(p)=0.5$ e $EI1(sol)=1.43$

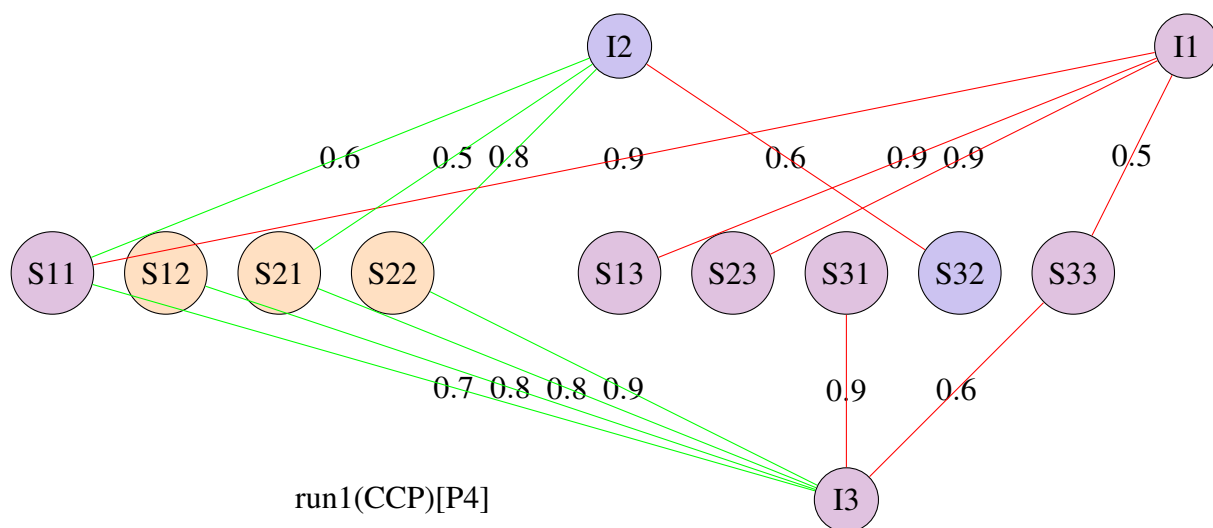


$I2(p)=0.6$ e $EI2(sol)=1.06$

Figura 17: Run1[P3]

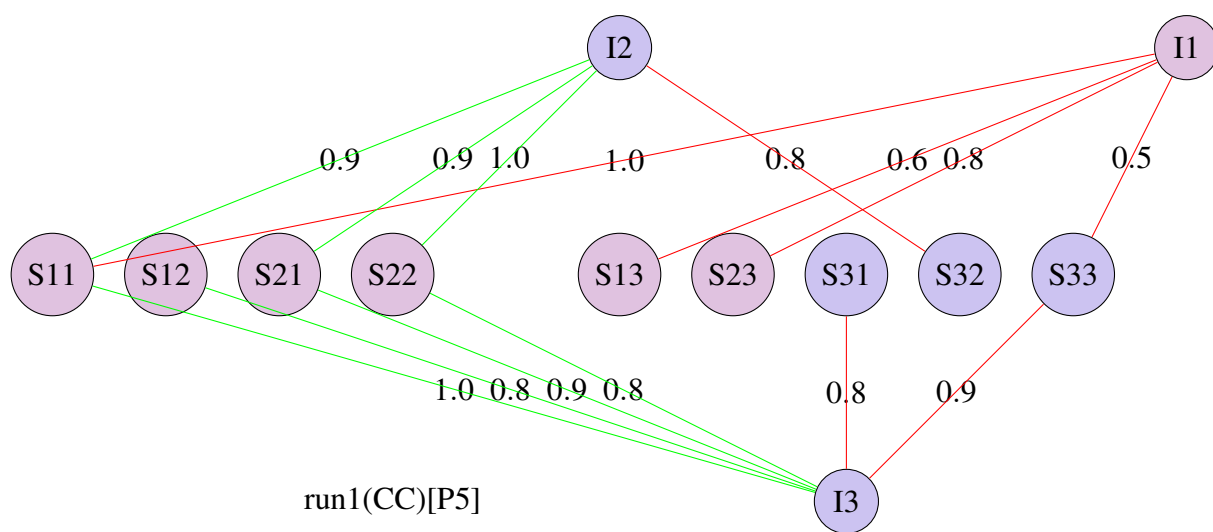


$I1(p)=0.5$ e $EI1(sol)=1.41$

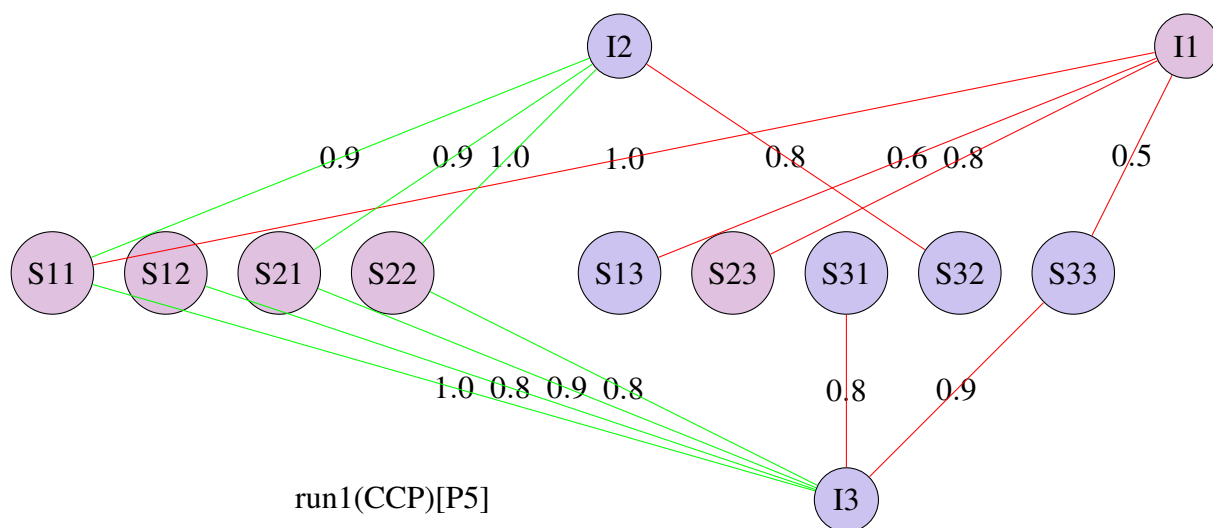


$I2(p)=0.7$ e $EI2(sol)=1.13$

Figura 18: Run1[P4]

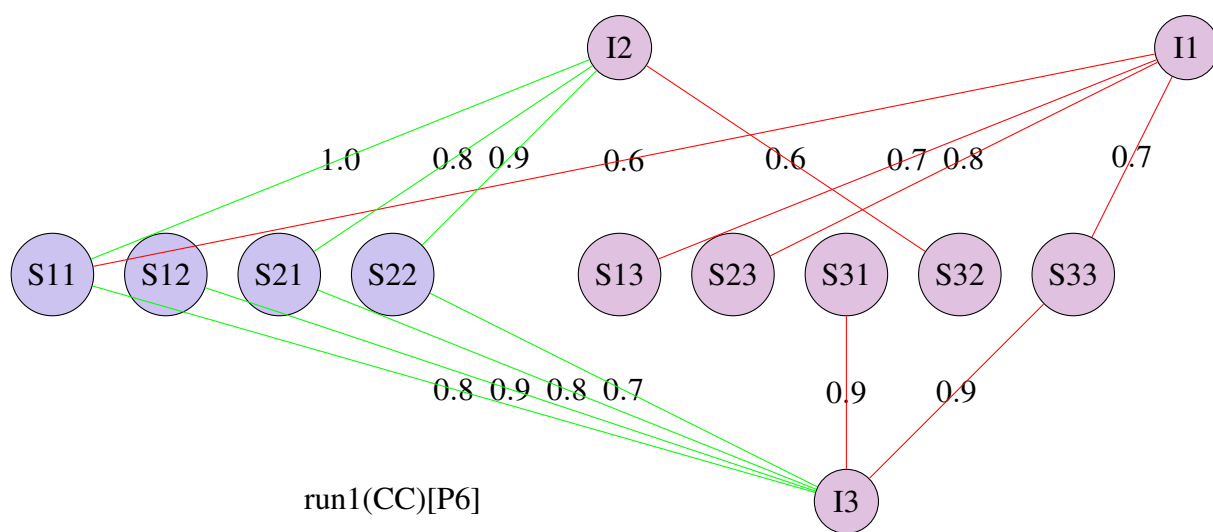


$I1(p)=0.5$ e $EI1(sol)=1.36$

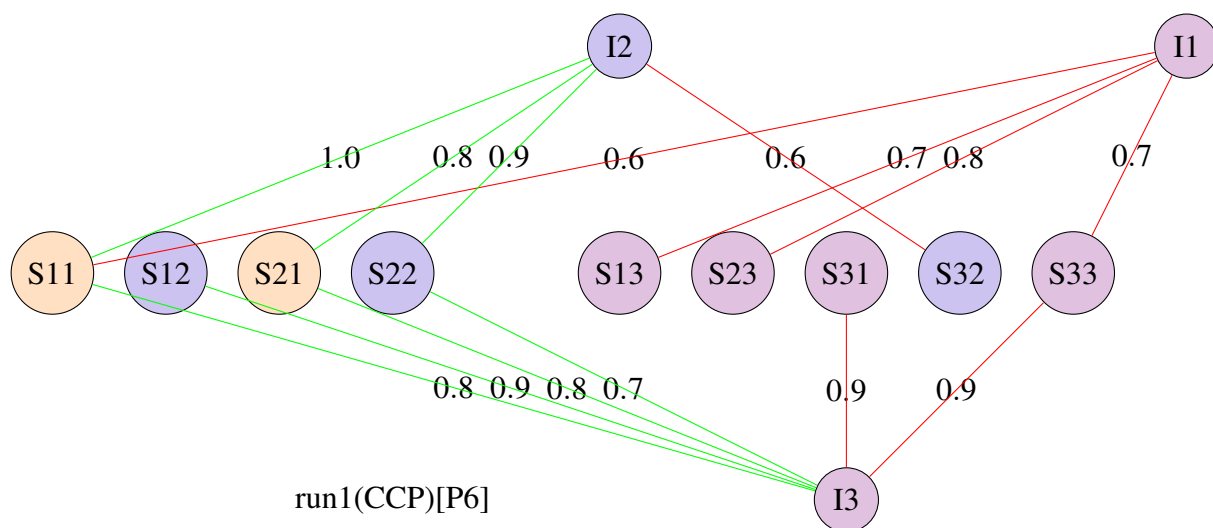


$I1(p)=1.1$ e $EI1(sol)=0.99$

Figura 19: Run1[P5]

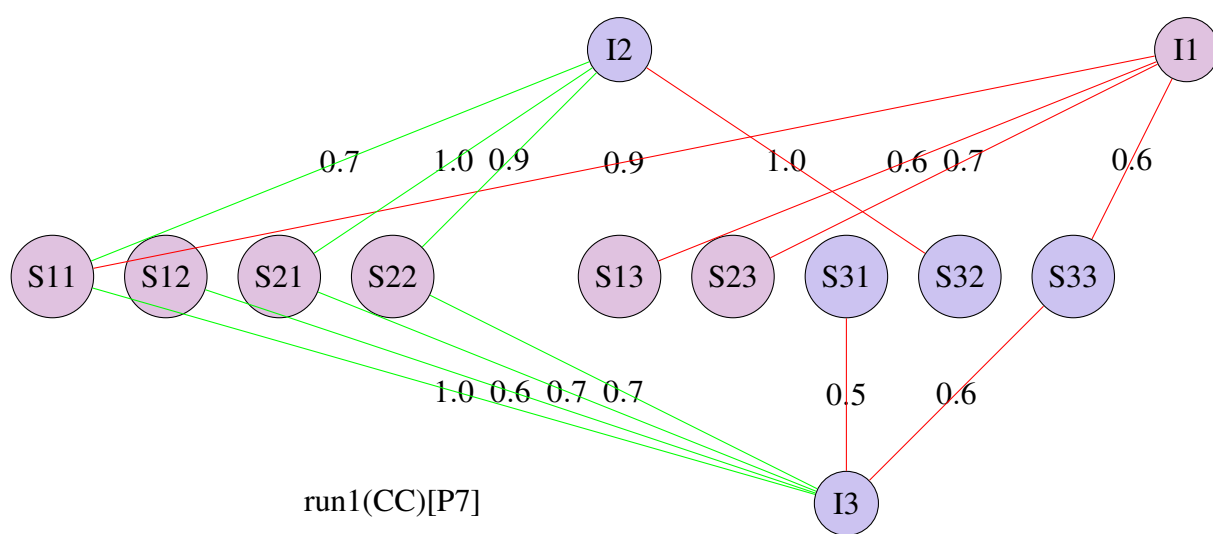


$I1(p)=0.6$ e $EI1(sol)=1.48$

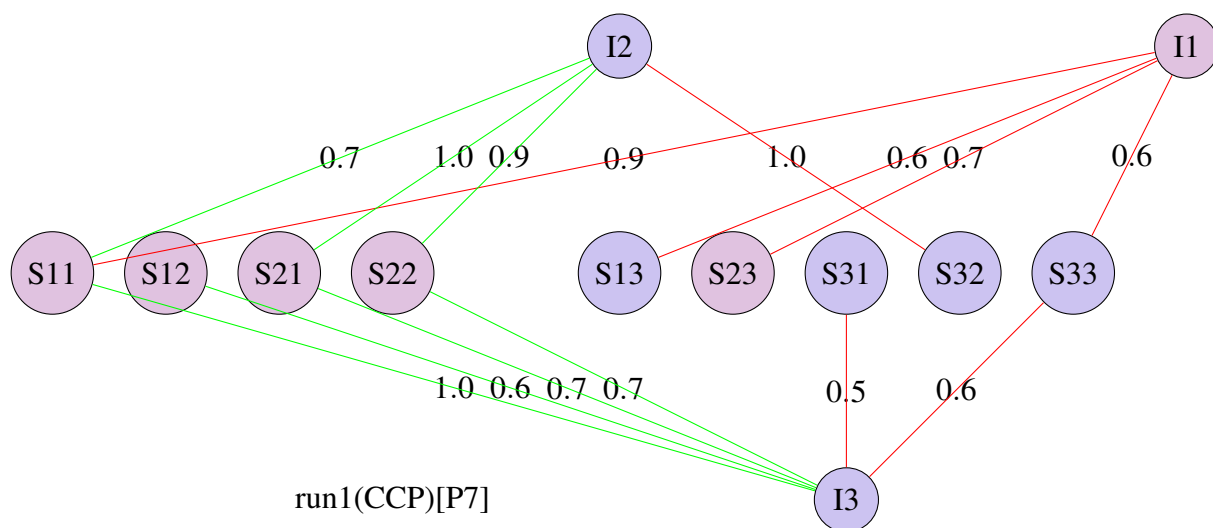


$I1(p)=0.7$ e $EI1(sol)=0.99$

Figura 20: Run1[P6]

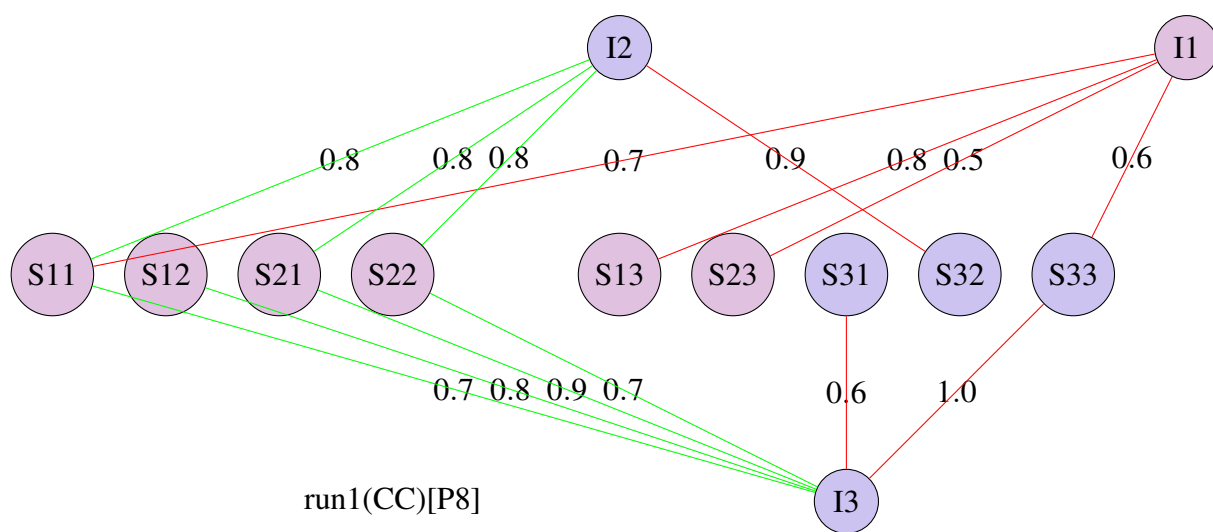


$$I1(p)=0.6 \text{ e } EI1(sol)=1.47$$

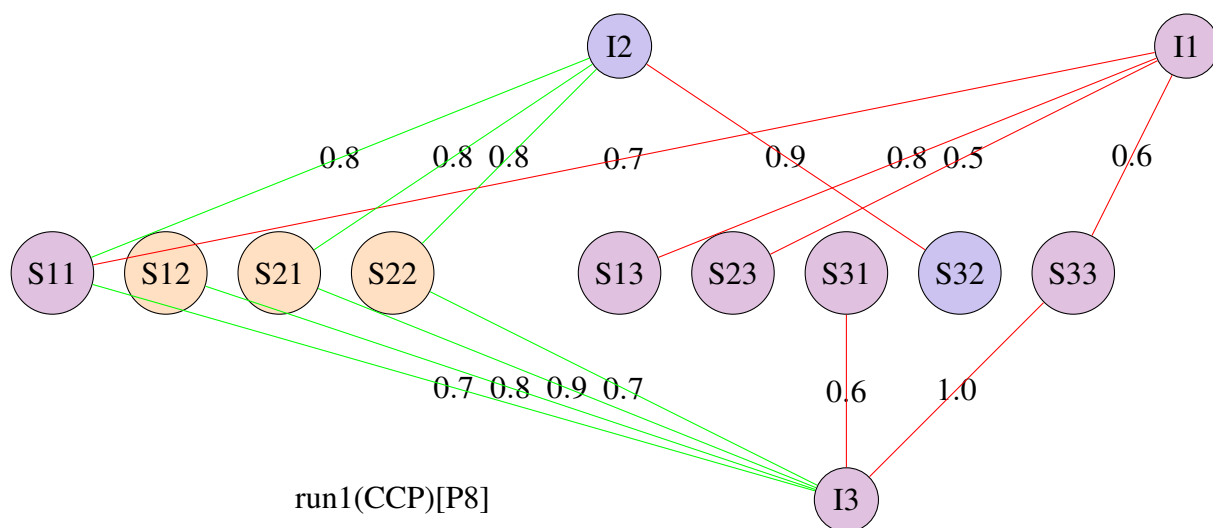


$$I1(p)=1.2 \text{ e } EI1(sol)=1.33$$

Figura 21: Run1[P7]

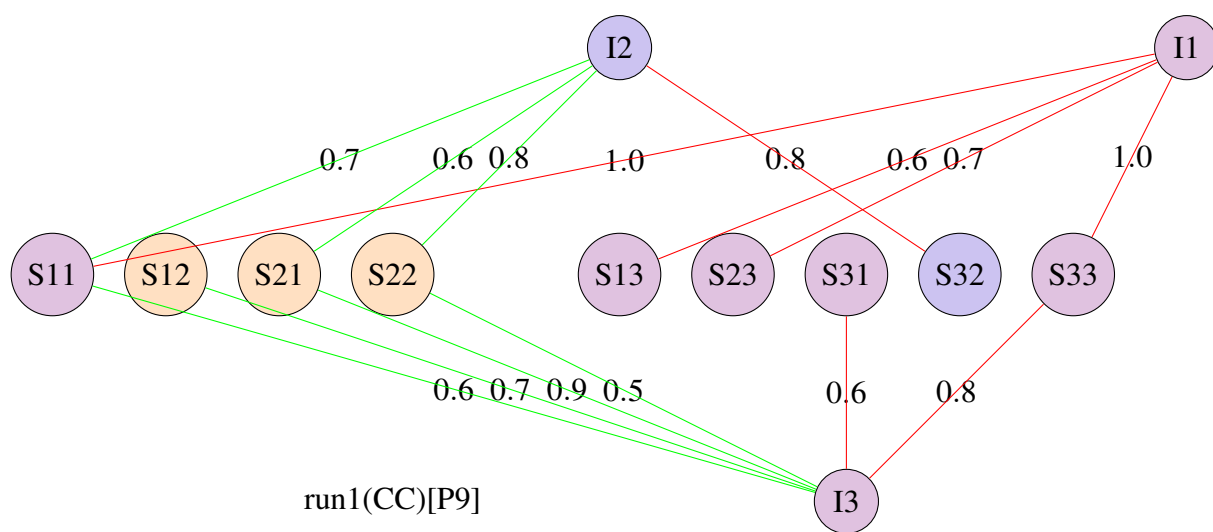


$I1(p)=0.6$ e $EI1(sol)=1.60$

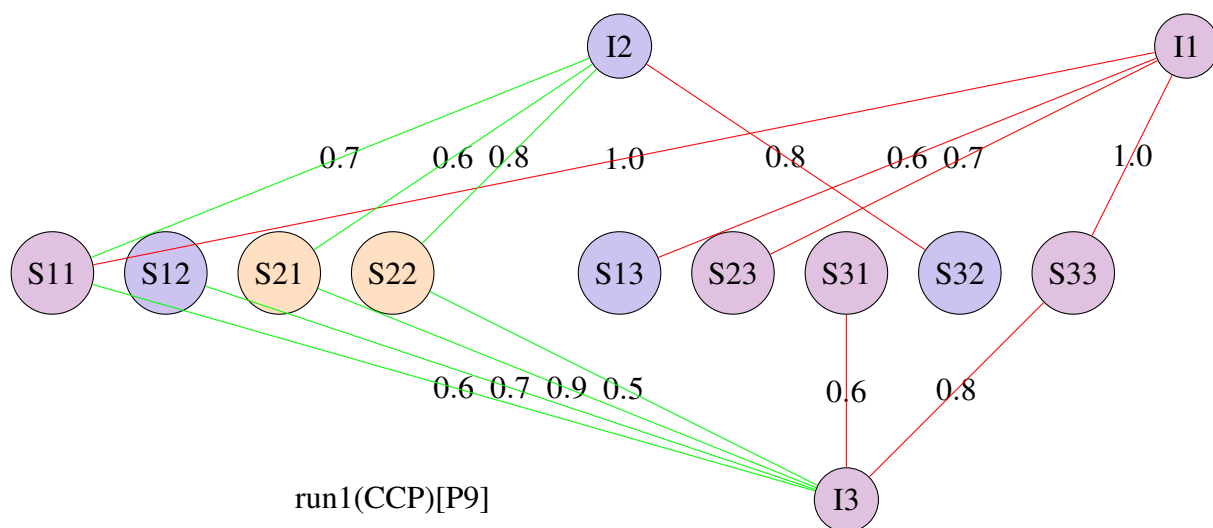


$I2(p)=0.7$ e $EI2(sol)=1.39$

Figura 22: Run1[P8]

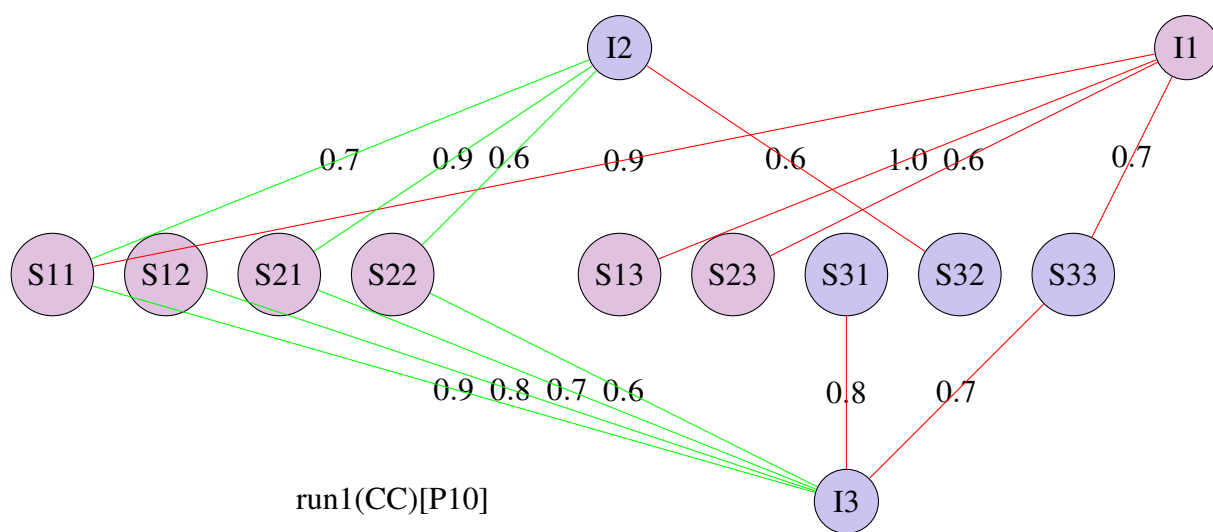


$I1(p)=0.6$ e $EI1(sol)=1.58$

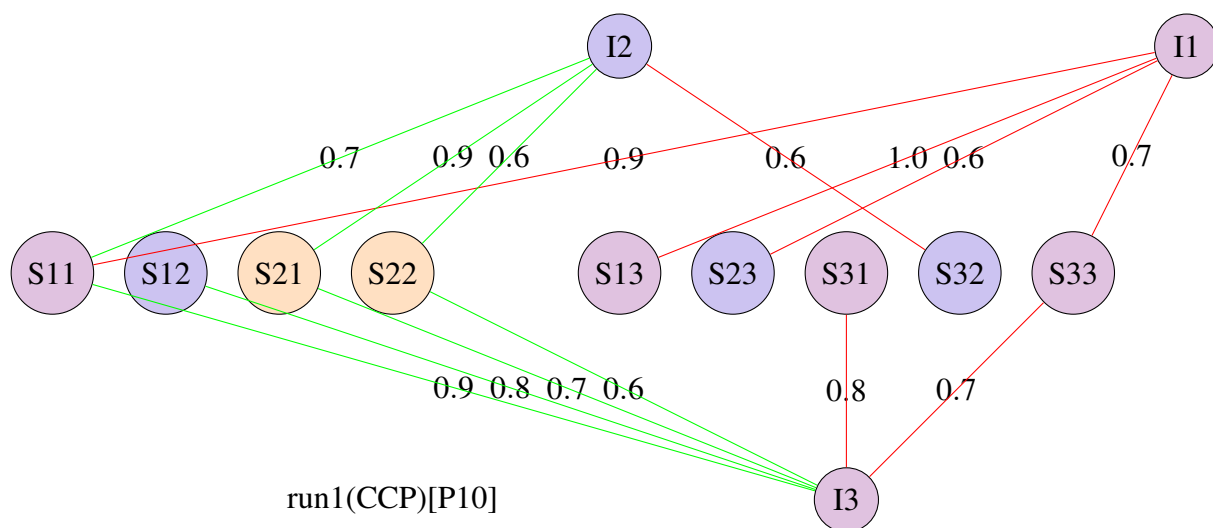


$I2(p)=1.2$ e $EI2(sol)=1.29$

Figura 23: Run1[P9]



$I1(p)=0.7$ e $EI1(sol)=1.98$



$I2(p)=1.5$ e $EI2(sol)=1.68$

Figura 24: Run1[P10]

5 Trabalhos futuros e conclusão

Em relação ao primeiro objetivo que era uma nova formulação para RCC, nossos resultados sugerem que para K pequenos o problema garante bons resultados computacionais em comparação ao modelo adotado na Literatura.

Quanto ao segundo objetivo, um modelo baseado em grafos pirobalísticos mostra uma boa possibilidade quando se une a ideia de agrupamento e probabilidade. Sabemos que, a corrupção vem crescendo nos últimos anos, no nosso país e em outras partes do planeta. Um modelo de Equilíbrio Estrutural têm um papel fundamental neste contexto, já que serve como ferramenta para de estudo de captura de empresas corruptas nas licitações públicas. Assim, o objetivo deste trabalho foi estudar algumas metodologias de agrupamentos para melhora do tempo computacional e para tentar prevê relações corruptas nos grafos que formam a rede de licitações. Utilizar a ideia da teoria de grafos de sinais probabilísticos foi uma ideia inovadora, criando conexão entre a teoria dos grafos de sinais e da inferência estatística. Decidir qual é a menor Perda Esperada, geram problemas de exponencias, mas trás uma perspectiva de todas as possibilidades de geração de uma rede dada.

O modelo foi utilizado para explorar rede de corrupção a nível micro e específico, pois o problema de geração de grafos é exponencial . Porém, como estamos interessados em um estudo local, se faz satisfatório, pois para uma rede que se tem uma visão global, perde-se muita informações e se torna mais difícil quantificar a corrupção.

Para trabalhos futuros, temos objetivo de fazer mais teste com RCC e trabalhar com grafos de sinais que possuem múltiplas arestas.

Por fim , acreditamos que o modelo apresentado para captura do Estado, embora possua limitações, contribui para o entendimento dos estudados de agrupamento em grafos para rede de licitações e pode ser estendidos para redes reais, na análise para ações no combate contra a corrupção.

Referências

- ABELL, P.; LUDWIG, M. Structural balance: a dynamic perspective. *Journal of Mathematical Sociology*, v. 33, pp. 129-155, 2009. Citado na página 12.
- ADEJUMO, G.; DUIMERING, P.; ZHONG, Z. A balance theory approach to group problem solving. *Social Networks*, vol. 30, pp. 83–99, 2008. Citado na página 12.
- ALVARO, J. L.; GARRIDO, A. *Psicologia social: perspectivas psicológicas e sociológicas*. São Paulo: McGraw-Hill, 2006. Citado na página 16.
- BANSAL, N.; BLUM, A.; CHAWLA, S. Correlation clustering. In: *Proceedings of the 43rd annual IEEE symposium of foundations of computer science*, pp. 238–250, 2002. Citado 2 vezes nas páginas 12 e 19.
- BRUSCO, M. et al. Two algorithms for relaxed structural balance partitioning: linking theory, models and data to understand social network phenomena. *Sociological Methods Research*, vol. 40, 57-87, 2011. Citado 2 vezes nas páginas 27 e 28.
- CAMPELO M., C. R. F. Y. Information processing letters. *Networks*, v. 89, pp. 1097-1111., 2004. Citado na página 22.
- CARTWRIGHT, D.; HARARY, F. A note on a matrix criterion for unique colorability of a signed graph. *Psychometrik*, v.32, pp. 291-296, 1967. Citado 2 vezes nas páginas 12 e 14.
- DASGUPTA, B. et al. Algorithmic and complexity results for decompositions of biological networks into monotone subsystems. *BioSystems*, vol. 90, pp. 161–178, 2007. Citado na página 12.
- DAVIS, J. A. Clustering and structural balance in graph signed. *Human Relations*, vol. 20, pp. 181-187, 1967. Citado na página 14.
- DOREIAN, P. A multiple indicator approach to blockmodeling signed networks. *Social Networks*, vol. 30, pp. 247-258, 2008. Citado 2 vezes nas páginas 19 e 28.
- DOREIAN, P.; MRVAR, A. A partitioning approach to structural balance. *Social Networks*, vol. 18, pp. 149-168, 1996. Citado 3 vezes nas páginas 12, 15 e 28.
- DOREIAN, P.; MRVAR, A. *Partitioning signed social networks*. [S.l.: s.n.], 2009. Citado 2 vezes nas páginas 12 e 19.
- DOREIAN, P.; MRVAR, A. Identifying fragments in networks for structural balance and tracking the levels of balance over time. *Connections*, v. 35, pp. 6-18, 2016. Citado na página 12.
- DRUMMOND L., F. R. F. Y.; LEVORATO, M. Efficient solution of the correlation clustering problem: An application to structural balance. *Springer Berlin Heidelberg. In Demey, Y. and Panetto, H., editors, on the Move to Meaningful Internet Systems: OTM 2013 Workshops Lecture Notes in Computer Science*, v. 8186 pages 674–683, 2013. Citado na página 12.

- ESTRADA, E.; BENZI, M. Walk-based measure of balance in signed networks: Detecting lack of balance in social networks. *Physical Review E*, v. 90, pp. 042802:1-10, 2014. Citado 4 vezes nas páginas 9, 12, 15 e 16.
- FAZEKAS, M.; TÓTH, I. J. *From corruption to state capture: A new analytical framework with empirical applications from Hungary*. [S.l.]: Government Transparency Institute, 2014. Citado na página 12.
- FAZEKAS, M.; TÓTH, I. J. Three indicators of institutionalised grand corruption using administrative data. *Corruption Research Centre*, 2014. Citado 4 vezes nas páginas 36, 37, 38 e 40.
- FAZEKAS, M.; TÓTH, I. J.; KING, L. P. *Anatomy of grand corruption: A composite corruption risk index based on objective data*. [S.l.]: Corruption Research Centre, 2013. Citado 3 vezes nas páginas 12, 35 e 36.
- FIGUEIREDO, R.; FROTA, Y. The maximum balanced subgraph of a signed graph: Applications and solution approaches. *European Journal of Operational Research*, vol. 236, pp. 473-487, 2014. Citado na página 12.
- FIGUEIREDO, R.; MOURA, G. Mixed integer programming formulations for clustering problems related to structural balance. *Social Networks*, v. 35, pp. 639-651, 2013. Citado 6 vezes nas páginas 9, 12, 19, 21, 22 e 25.
- FROTA Y., M. N. N. T. R. C. A branch-and-cut algorithm for partition coloring. *Networks*, v. 55, pp. 194-204., 2010. Citado 2 vezes nas páginas 12 e 22.
- GAREY M.R., J. D. Computers and intractability. *W.H. Free man and Company, New York.*, v. 24th ed, 2003. Citado na página 21.
- GOUNEV, P.; BEZLOV, T. Examining the links between organised crime and corruption. *Sofia: Center for the Study of Democracy*, 2010. Citado na página 35.
- HARVEY, J. H. Fritz heider (1896–1988). *American Psychologist*, vol. 44, 570-571, 1989. Citado na página 12.
- HEIDER, F. Attitudes and cognitive organization. *Journal of Psychology*, vol. 21, pp. 107-112., 1946. Citado 2 vezes nas páginas 12 e 14.
- HELLMAN, J. S. et al. Measuring governance and state capture: The role of bureaucrats and firms in shaping the business environment. In: *World Bank Working Paper 2312*, 2000. Citado na página 12.
- HUFFNER, F.; BETZLER, N.; NIEDERMEIER, R. Optimal edge deletions for signed graph balancing. In: *Proceedings of 6th WEA in LNCS*, vol. 4525, 297–310, 2007. Citado na página 12.
- LEMANN T.B., S. R. Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry*, vol. 15, pp. 7–90, 1952. Citado na página 27.
- LEVORATO, J. *Efficient solutions to the correlation clustering problem*. Dissertação de Mestrado (UFF), 91 fls, Niterói, Rio de Janeiro: [s.n.], 2015. Citado 3 vezes nas páginas 9, 20 e 22.

- MCKINNEY, J. C. An educational application of a two-dimensional sociometric test. *Sociometry*, v.11, pp. 356-367, 1948. Citado na página 27.
- MENDONÇA, I. et al. Relevance of negative links in graph partitioning: A case study using votes from the european parliament. In: *2nd European Network Intelligence Conference (ENIC)*, 2015. Citado na página 12.
- NEWCOMB, T. *The acquaintance process*. [S.l.]: Holt, Reinhard Winston, 1961. Citado na página 28.
- RAMOS, C. D. Corrupção e o papel do estado: uma análise sobre o controle e a intervenção estatal. *Revista Banco de Idéias: Encarte - Revista* 37, Instituto Liberal, vol. 37, pp. 1-29, 2013. Citado na página 35.
- SAMPSON, S. *A novitiate in a period of change: An experimental case study of relationships*. [S.l.: s.n.], 2009. Citado na página 26.
- TRAAG, V. A.; BRUGGEMAN, J. Community detection in networks with positive and negative links. *Physical Review E*, vol. 80, pp. 036115:1-6, 2009. Citado na página 12.
- TRICK, M. A. A column generation approach for graph coloring. *INFORMS Journal of Computing*, vol.8, pp. 334-354, 1996. Citado na página 19.
- WOLSEY, L. *Integer Programming*. [S.l.]: Wiley, 1998. Citado 2 vezes nas páginas 23 e 24.
