

Ciência e Visualização de Dados em Saúde

MO413/MC936 - 2024/1

Relatório do trabalho T3 - Análise de Enriquecimento de Conjuntos de Genes (GSEA) e Análise de Rede das vias enriquecidas

Equipe Get Out Zika

Nome	RA	Especialização
Bruno Brito Pereira da Silva	b213947	Biologia
Márcio Silva Cruz	m290232	Computação
Rafael Simionato	r066505	Computação

Introdução

O sequenciamento genético de nova geração, principalmente o de RNA, ajudou a entender de maneira mais profunda como os genes são ativados e inativados em diferentes estados nos organismos, seja quando o corpo esteja saudável ou em desequilíbrio em estado patológico. Partindo desse princípio, estuda o sequenciamento de RNA em pacientes com artrite reumatoide ajuda a entender quais genes estão “*down*” regulados e quais estão “*up*” regulados. O que por fim, permite que sejam desenvolvidos novos tratamentos específicos para esta doença autoimune que afeta 2 vezes mais mulheres do que homens.

Metodologia

Abaixo estão organizados os passos executados para a realização deste trabalho.

Geração da rede em código Python

A primeira parte deste exercício compreendeu a geração de uma rede bipartida sendo a mesma gerada por nós representando vias de expressão gênica, e nós representando genes possivelmente associados a tais vias. As arestas que ligam os nós representam a relação entre uma via e um gene.

Essa rede foi gerada a partir de um processo chamado análise de enriquecimento de enriquecimento. Dada uma tabela contendo dados de expressão diferencial comparando dois grupos de pacientes com e sem artrite reumatóide, aplicou-se o método de enriquecimento GSEA buscando se identificar em sua base de dados quais genes estão

associados a quais vias de expressão. Tal método gera um ranqueamento com base nos genes e níveis de expressão informados como entrada e por fim, estabelece um valor de significância estatística para cada resultado obtido.

De posse dessas informações, foi possível criar uma rede informando como nó origem uma via de expressão e como nó destino um gene associado a essa via, sendo que este processo se repete para todo gene associado a referida via. A rede final gerada foi exportada para ser analisada através do aplicativo Cytoscape.

Análise de centralidade

Após a importação da rede gerada para o Cytoscape, procedeu-se a análise topológica, onde foram geradas diversas medidas para avaliação. A Figura 1 apresenta a rede de vias enriquecidas, expressando o tamanho de seus nós em função da medida de centralidade *betweenness centrality*.

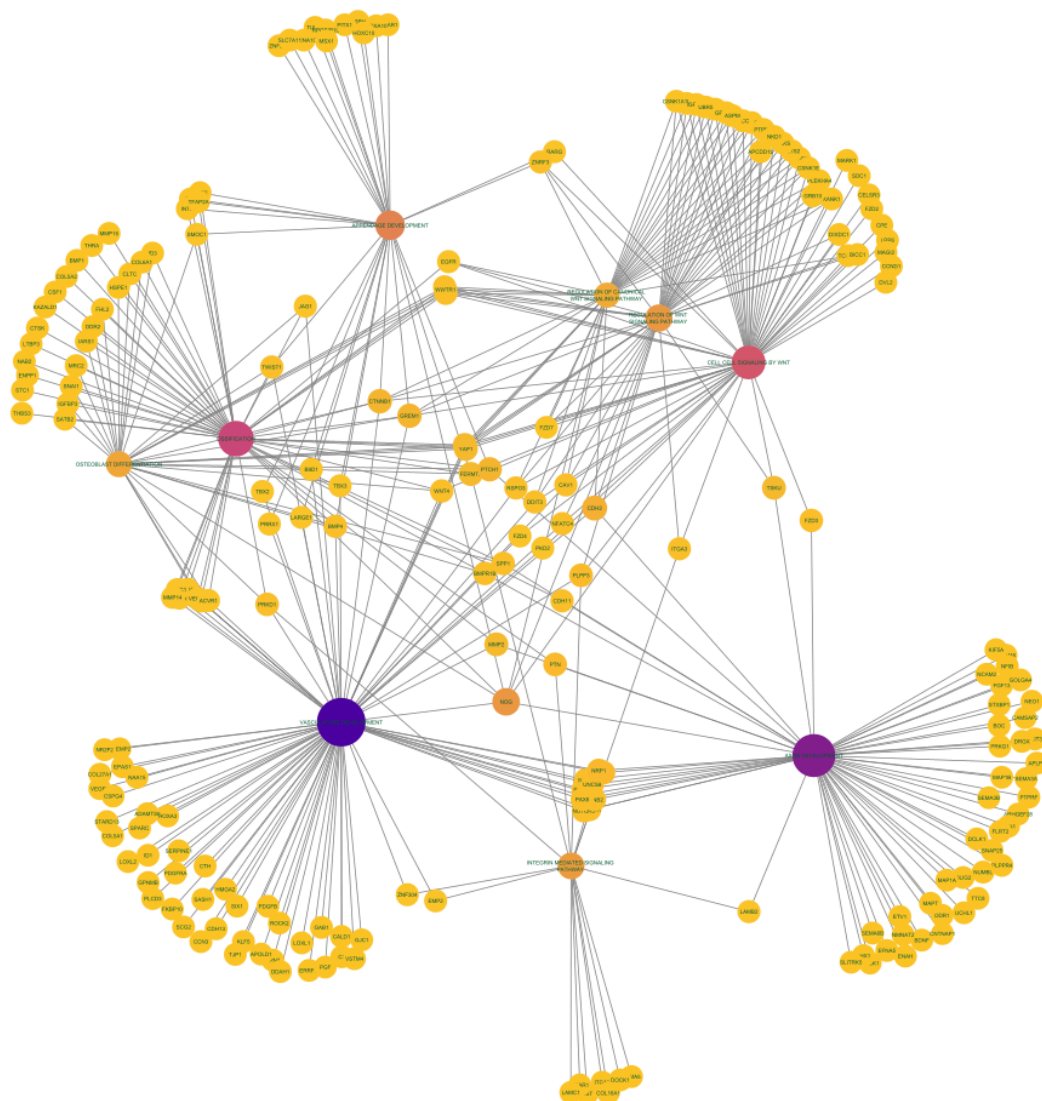



Figura 1: Visão geral da rede de vias enriquecidas

Além do tamanho do nó, para expressar sua centralidade, também foi utilizada a estratégia de coloração, onde os nós representados pelas cores mais escuras demonstram maior centralidade, e os nós com coloração amarela apresentam menor centralidade. A Tabela 1 evidencia os nós e suas respectivas centralidades; organizadas da maior para a menor.

Tabela 1: relação de nós e centralidades calculadas

	label	BetweennessCentrality
	VASCULATURE DEVELOPMENT	0.4515793160619104
	AXON DEVELOPMENTZ	0.354429723587146
	OSSIFICATION	0.22473168384030948
	CELL CELL SIGNALING BY WNT	0.19269162777199084
	APPENDAGE DEVELOPMENT	0.11760513614550223
	NOG	0.08290273087178483
	REGULATION OF WNT SIGNALING PATHWAY	0.07703015118023408
	INTEGRIN MEDIATED SIGNALING PATHWAY	0.06905134484599301
	OSTEOBLAST DIFFERENTIATION	0.05646695671666619
	REGULATION OF CANONICAL WNT SIGNALING PATH...	0.044133358095861615
	CDH2	0.038632105222900885
	CTNNB1	0.027780348445981665
	GREM1	0.027780348445981665
	PTCH1	0.025749906951548655

É possível observar pela análise dos dados, na tabela, que as duas vias mais importantes são **VASCULATURE DEVELOPMENT** e **AXON DEVELOPMENTZ**, respectivamente com centralidade calculada de 0.4515793160619104 e 0.354429723587146. A Figura 2 mostra, essa região da rede, onde é possível perceber a alta quantidade de arestas ligando diversos genes distintos e em comum a essas duas vias.

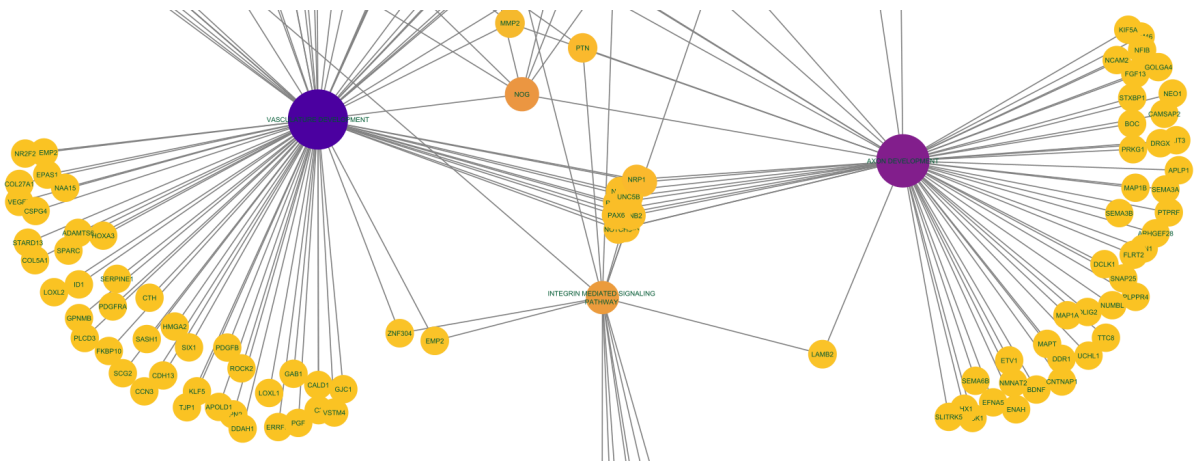


Figura 2: Vias mais importantes pelo critério de *betweenness centrality*.

Utilizando-se a mesma técnica é possível identificar os genes hubs que apresentam maior centralidade. As 10 primeiras linhas da tabela 1 (com exceção da sexta linha - NOG) representam vias de expressão, sendo que em seguida estão as linhas que representam os

genes. Assim, uma vez que a tabela está organizada por *betweenness centrality*, os genes com maior centralidade são: **NOG**, **CDH2**, **CTNNB1** e **GREM1**, respectivamente com centralidade calculada de 0.08290273087178483, 0.038632105222900885, 0.027780348445981665 e 0.027780348445981665. Curiosamente os genes CTNNB1 e GREM1 apresentam mesmo grau (7), centralidade calculada (0.0277) e se ligam às mesmas 7 vias. A Figura 3 evidencia em vermelho as ligações entre os genes hubs mencionados acima e as vias às quais se relacionam.

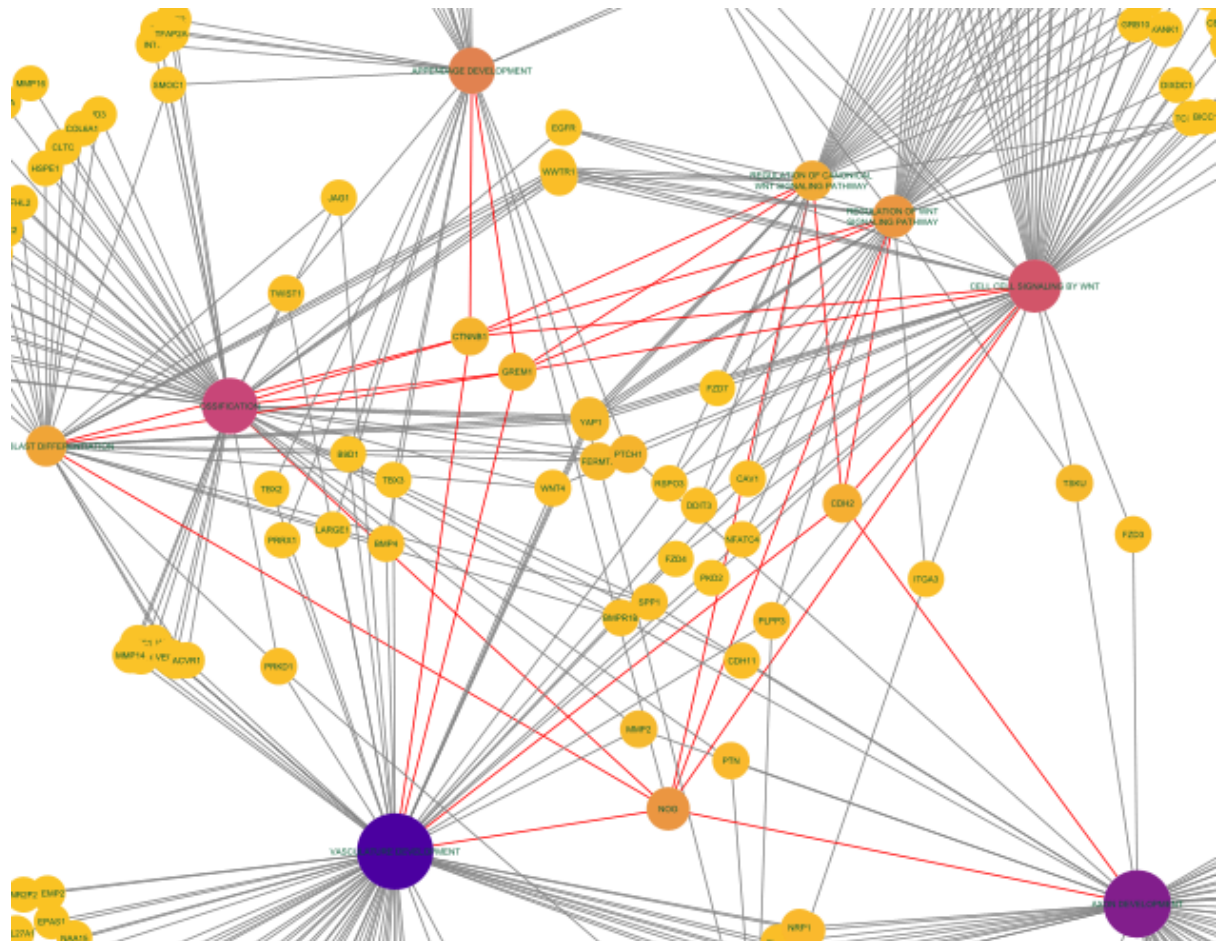


Figura 3: Evidenciamento dos genes hubs e vias de expressão relacionadas

Análise de comunidades

Para estudar a detecção de comunidades, foram utilizados 3 algoritmos (CyFinder, Louvain e OSLOM) descritos a seguir:

CyFinder: Usado para encontrar vários tipos de sub-redes com o objetivo de identificar biomarcadores em redes de doenças. Suas principais características são:

- 1 - Encontrando Subgrafos de Clique.
- 2 - Encontrando Cliques Máximos.
- 3 - Encontrando sub-grafos bipartidos.

- 4 - Encontrando Bicliques Máximos de Subgrafo Bipartido.
- 5 - Layout bipartido para visualização de conjunto partido de 2 colunas.
- 6 - Encontrando Componentes Conectados.
- 7 - Detecção de comunidade com múltiplos algoritmos.
- 8 - Encontrando Floresta de Extensão Mínima/Máxima
- 9 - Intersecção de Redes e Coleções como forma de análise de similaridade.
- 10 - Encontrando Crosstalk entre Subgrafos disjuntos.

A Figura 4 apresenta as comunidades encontradas através do CyFinder no Cytoscape.

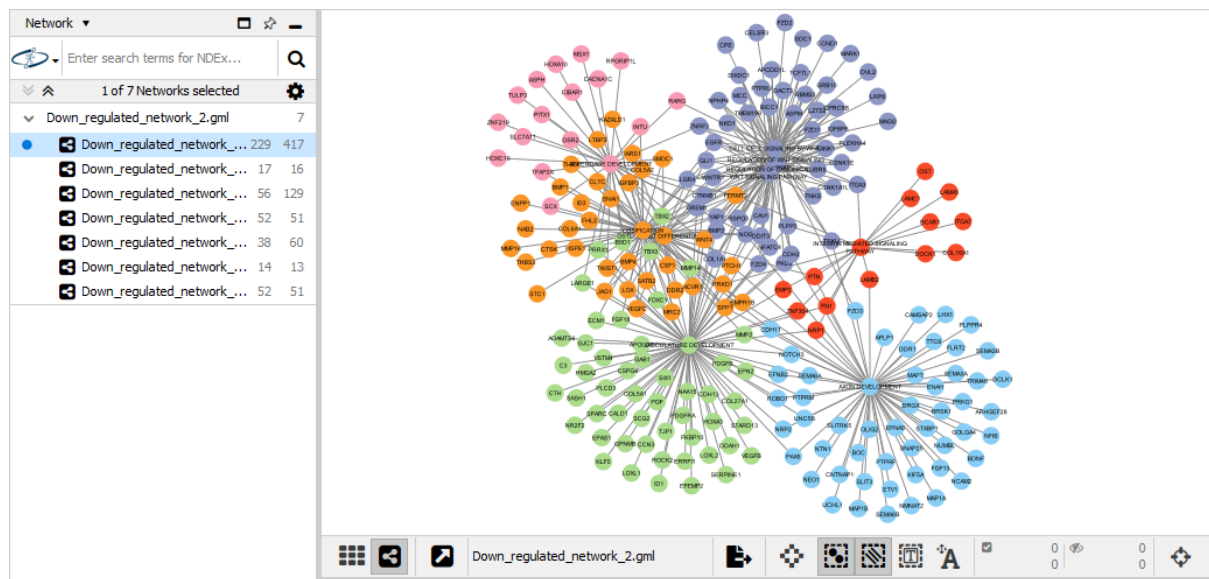


Figura 4: CyFinder - 1º algoritmo

Louvain: O algoritmo considera cada nó uma comunidade, e ao agrupar os nós vizinhos é calculada a modularidade, criando assim novas comunidades. Cada comunidade será vista como um nó e a partir deste nível irá repetir essa dinâmica até esgotar as possibilidades. Em resumo, são comunidades que se juntam a comunidades maiores (grupos que se juntam em grupos), sendo que, cada nó será um cluster onde irá conter um vetor de todos os nós que fazem parte daquela comunidade. A Figura 5 apresenta as comunidades encontradas através do Louvain no Cytoscape.

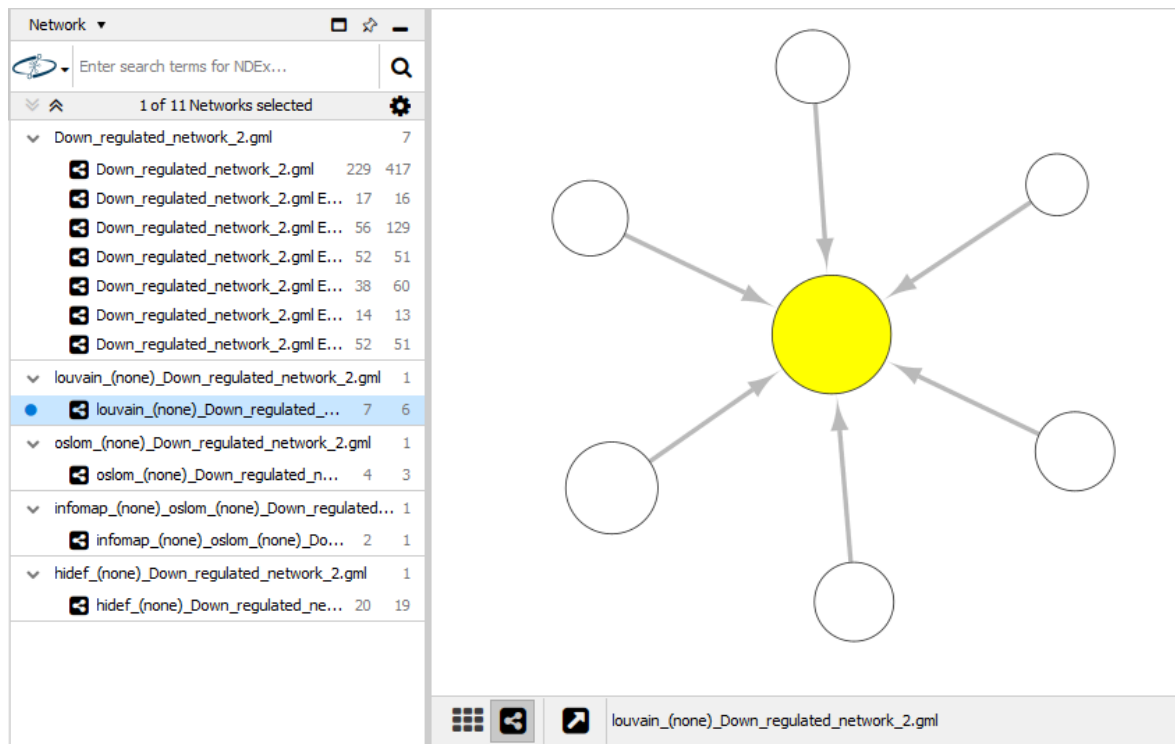


Figura 5: Louvain - 2º algoritmo

OSLOM: Similar ao Louvain, porém criam-se comunidades hierarquicamente. A Figura 6 apresenta as comunidades encontradas através do OSLOM no Cytoscape.

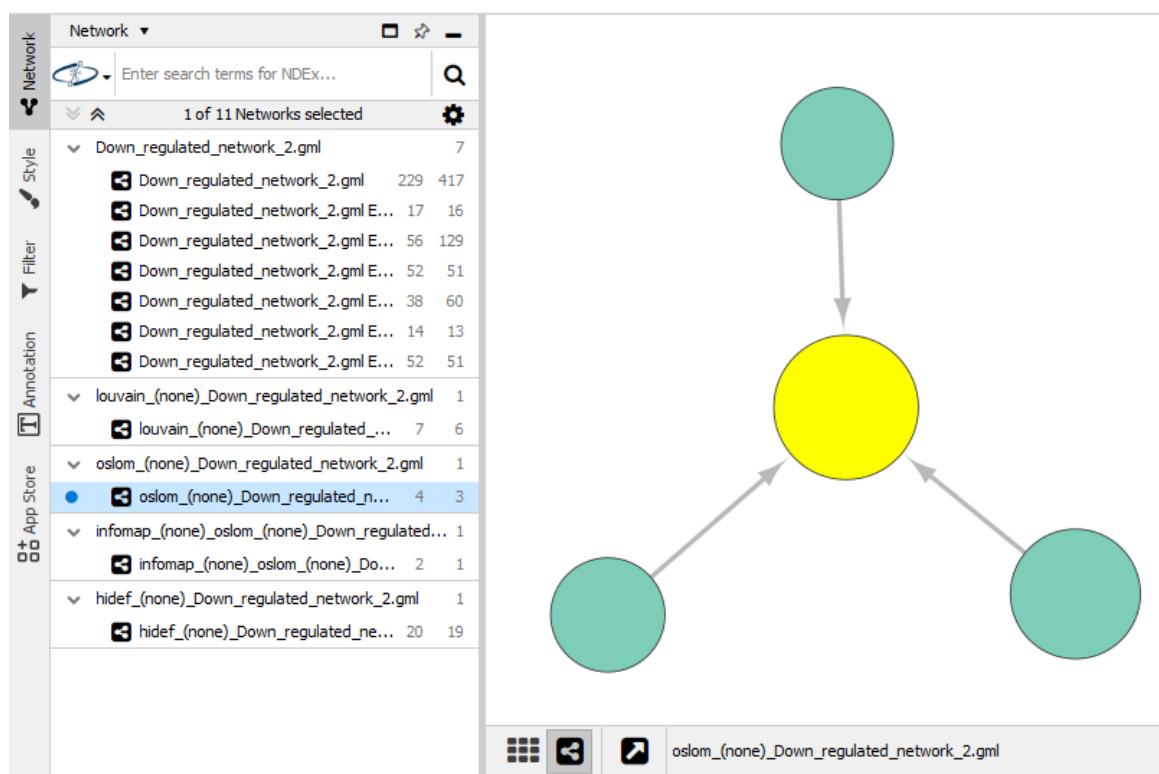


Figura 6: OSLOM - 3º algoritmo

Visão geral das comunidades detectadas

Uma visão geral das comunidades encontradas através dos algoritmos supracitados, é apresentada na Figura 7. Também consta nesta figura o resultado da execução de outros dois algoritmos: Infomap e Hídef.

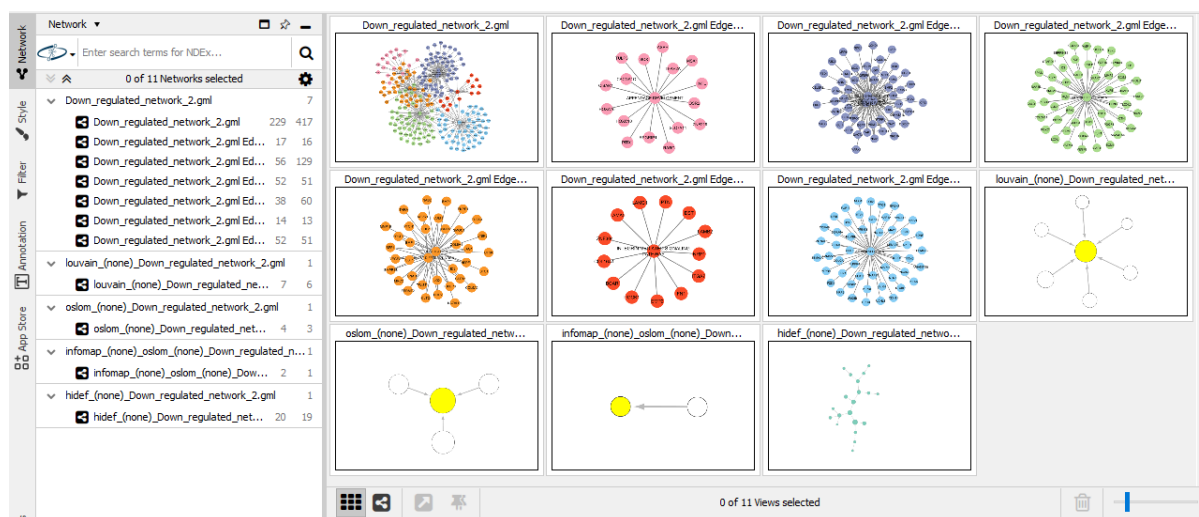


Figura 7 - Visão geral das comunidades

Interpretação biológica

Na análise das redes verificamos a existência de uma comunidade de genes que interagem com os hubs, os genes em questão são NOG, CDH2, CTNNB1 e GREM1. Esses genes em si atuam em sua maioria na formação e diferenciação de osso e de cartilagem (NOG e CDH2) e na interação célula-célula (CTNNB1) já GREM-1 por ser um antagonista de BMP (Bone Morphogenetic Protein) teoricamente já deveria ter uma sub-expressão e no caso, ela deveria apresentar uma expressão maior, pois ela inativa genes responsáveis por ossificação como o BMP-1. No entanto, estudos apontam que a inibição de GREM-1 melhora a osteoartrite em camundongos, e isso leva a entender que a sub-expressão de GREM-1 em humanos não é suficiente para inativar a degeneração de cartilagem. Neste contexto, provavelmente a osteoartrite necessita de uma inativação completa de GREM-1 para melhorar a evolução da doença.

Tendo em vista a atividade supramencionada para o gene GREM-1, analisamos a atividade de BMPs na rede e verificamos que todas as BMPs estão “down” reguladas. O que mostra que a baixa atividade de BMP atrelado a baixa de atividade de GREM-1 continua afetando a ossificação. No geral, conseguimos perceber que vários genes relacionados à formação e diferenciação de tecido ósseo estão com uma regulação menor, o que era previsível, mas ao nos depararmos com genes que são antagonistas das vias de produção de tecido ósseo “down” reguladas, fomos surpreendidos, pois era esperado que esta via estivesse “up” regulada, o que nos leva a entender que não é somente um gene que será responsável por toda desordem que um organismo vai passar a ter.