

## Сети заболеваний человека и проблема старения

Д.И. Чернышова, И.А. Евин, А.В. Мелерзанов

Московский физико-технический институт (государственный университет)

Ключевые слова: безмасштабные сети, интерактом, модуль заболевания

В конце двадцатого века на основе теории графов сформировалась новая область статистической физики – теория сложных сетей, ставшая эффективным инструментом исследования сложных систем различной природы, в том числе биологии и медицины [1]. В последние годы приложения теории сложных сетей к проблемам возникновения болезней человека привело к возникновению нового направления в медицине – сетевой медицины (NetworkMedicine)[5-11]. Цель данной статьи – представить краткий обзор наиболее важных публикаций этого научного направления, прежде всего связанных с пониманием проблемы взаимосвязи различных заболеваний.

В современной биологии и медицине значительные усилия направлены на нахождение связей между молекулярно-генетическим происхождением заболевания и его фенотипическим проявлением в виде симптомов. Хотя часто заболевания лечатся независимо от других, мало кто сомневается, что болезни связаны между собой. В 2007 году была построена первая сетевая структура заболеваний человека, в которой каждому узлу соответствует определенное заболевание и между узлами существует связь, если соответствующие им заболевания вызваны каким-то одним генетическим изменением[5]. Позднее аналогичные сетевые структуры стали создавать на основе патологий в метаболических реакциях, белковых взаимодействиях, регуляторных и сигнальных сетях и т.д., а также учитывались взаимодействия людей в социальных сетях, что позволило начать изучение проблем медицины методами теории многослойных сетей [2].

Чаще всего болезнь вызвана аномалиями при взаимодействии процессов на молекулярном уровне. Взаимосвязи между этими процессами отображаются в интерактоме – многослойной сетевой структуре включающей в себя все взаимодействия в живой клетке: от сети взаимодействий белок-белок до регуляторных сетей и сетей метаболических реакций. Несколько лет назад появились основания для предположения, что участвующие в аномальных взаимодействиях белки образуют в интерактоме кластер получивший название модуль заболевания.

## Сети заболеваний человека

Современная система классификации заболеваний человека была основана в конце девятнадцатого века из наблюдений корреляций между анализом патологий и клиническими синдромами и в значительной степени зависела от искусства определения синдромного фенотипа. В течение двадцатого века эта процедура стала более объективной, поскольку стала опираться еще и на молекулярно-биологические диагностические тесты. Тем не менее, эта классическая диагностическая схема имеет недостатки, признаваемые многими специалистами как из-за отсутствия точности в выявлении доклинической стадии заболевания, так из-за отсутствия специфичности при попытках определить заболевание однозначно. Однако, последние успехи молекулярной биологии дают возможность современной медицине диагностировать заболевания гораздо более точно.

В качестве примера можно привести недавние исследования, которые показали, что если у человека одна из аллелей содержит ген *FTO*, то у этого человека риск ожирения возрастает на 30% , а если этот ген содержится в обеих аллелях, риск ожирения возрастает до 67%. Такая связь между геном *FTO* и ожирением может служить примером сильной прямой связи между генотипом и фенотипом. При этом избыточный вес оказывается тесно связанным такими заболеваниями, как сахарный диабет, астма и другими.

Недавно Кристакис и Фаулер изучив данные многолетних медицинских наблюдений различных социальных групп в одном из районов штата Массачусетс (США), обнаружили, что социальные связи различной природы оказывают на возникновение избыточного веса не меньшее влияние, чем генетические факторы. Например, вероятность того, что если друг вашего друга достаточно хорошо знаком с человеком с избыточным весом, он сам наберет избыточный вес, на 20% выше, чем в социальных сетях со случайными связями между людьми. Было показано, что ожирение подобно инфекциям распространяется по социальным сетям.

В настоящее время имеется множество данных, в которых устанавливаются связи между заболеваниями в виде сетей. Например, Rzhetsky *et al* установили связи между 161 сопутствующими друг другу заболеваниями на основе данных взятых из истории болезней полутора миллиона пациентов по программе Medicare[9]. Hidalgo *et al* построили сеть фенотипических признаков заболеваний используя данные о сопутствующих друг другу болезнях более 30 миллионов пациентов [6], показав, симптомы не только по-прежнему являются решающим фактором правильного клинического диагноза и успешного лечения, но и важнейшим ресурсом для теоретического анализа.

### **Выявление заболеваний из неполного интерактома**

Несомненно, важно и необходимо уметь выявлять мутации, ведущие к заболеваниям, в контексте интерактома человека – сетевая структура, всесторонне отражающей все биологически значимые молекулярные взаимодействия. Однако, предсказательные возможности современных подходов, основанных на теории сложных сетей, ограничены множеством причин. Во-первых, сегодня нам известно лишь около 20% общей структуры интерактома и поэтому в поисках молекулярных механизмов заболеваний вынуждены опираться на пока еще очень неполные данные о сети белковых взаимодействий. Во-вторых, как уже упоминалось выше, продукты мутаций генов – соответствующие белки – рассеяны по интерактому не случайным образом, а имеют тенденцию взаимодействовать друг с другом, формируя один из нескольких подграфов, называемым модулем заболевания (disease module).

В работе [7] даны примеры взаимодействий модулей заболеваний для рассеянного склероза (multiple sclerosis(MS)), пароксизмальные расстройства (peroxisomal disorders(PD)) и ревматоидного артрита (rheumatoid arthritis (RA)). Заболевания с перекрывающимися модулями (MS и RA) имеют фенотипические сходства и высокую вероятность быть сопутствующими заболеваниями. Не пересекающиеся модули заболеваний, такие как MS и PD, не имеют выявленных общих клинических симптомов.

Однако, несмотря на частое использование или просто упоминания в научных дискуссиях и докладах, гипотеза о модулях заболевания не имеет конкретных алгоритмов и правил их выявления и построения, а также сколько-нибудь надежного обоснования с точки зрения современной теории сложных сетей.

### **Фрагментация модулей заболеваний**

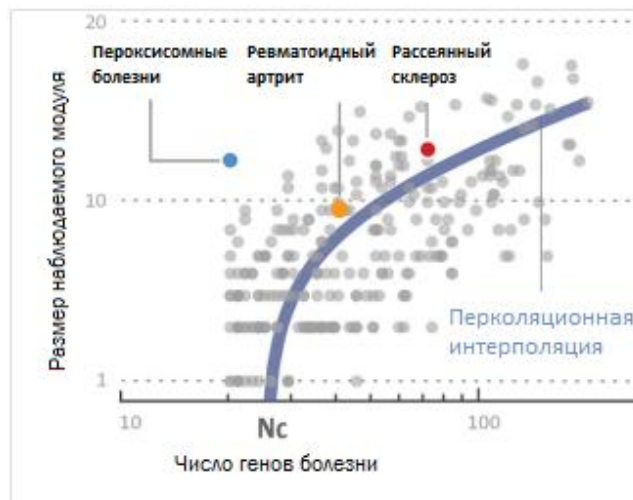
Построение интерактома было начато с составления сети из 141296 физических взаимодействий между 13460 белками, экспериментально задокументированных в клетках человека. Сюда вошли белок-белковые, регуляторные взаимодействия и метаболические реакции. Также был составлен список 299 заболеваний из медицинского рубрикатора заболеваний (Medical Subject Headings (MeSH)) содержащий не меньше 20 связанных генов на основе данных, содержащихся в Mendelian Inheritance in Man (OMIM) и баз данных genome-wide association study (GWAS), включающих в себя 2436 белков, связанных с болезнями. Несмотря на некоторый прогресс, значительная часть интерактома из списка болезней, связанных с заболеваниями генов остается несвязанной.

Например, для рассеянного склероза из 69 белков всего 11 образуют связанный подграф, остальные же случайным образом рассеяны по интерактому. Эта картина имеет место для всех 299 заболеваний. Их выявленные модули составляют около 20% соответствующих генов болезни.

Существует несколько причин фрагментации интерактома как сетевой структуры. Основной причиной является недостаточность данных: отсутствующие связи оставляют многие

связанные с заболеваниями белки изолированными в их модулях заболеваний. Согласно важнейшему разделу теории сложных сетей - теории перколяции, если имеется определенная доля  $p$  связей, связанный подграф (в данном случае модуль заболевания) содержащий  $m$  узлов, при значения  $p$  выше критического значения испытывает фазовый переход при котором формирует наблюдаемый модуль. Если же это значение ниже критического, модуль остается слишком фрагментированным и не воспринимается как нечто целое.

Детальные вычисления показали, что согласно теории перколяции, для заболевания причиной которого было меньше 25 генов, его модуль слишком фрагментирован, чтобы быть наблюдаемым в интерактоме. То есть, только заболевания, причиной которых было больше 25 генов, имеют хорошо наблюдаемые модули заболеваний (Рисунок 1).



**Рисунок 1.** Оценка критического размера модуля заболевания на основе теории перколяции.

### Фенотипическая Сеть Болезней (Phenotypic Disease Network – PDN)

Хотя успехи в использовании сетевых методов на молекулярно-генетическом уровне для изучения различных заболеваний весьма впечатляющи, немало полезной информации можно получить применяя сетевые технологии при анализе фактических данных по заболеваниям на фенотипическом уровне в виде истории болезней пациентов, содержащих информацию о связях различных заболеваний между собой и прогрессе этих заболеваний.

Чтобы понять связи болезней, каждую из них определяют как определенный специфический набор фенотипов влияющих на одну или несколько функциональных систем. При этом подразумевается, что живой организм может быть описана в сложном фенотипическом пространстве как связанная из отдельных элементов дискретная система. Для этого в 2009 году была построена Фенотипическая Сеть Болезней (Phenotypic Disease Network (PDN)) на основе изучения корреляций историй болезней более тридцати миллионов пациентов [6].

Было установлено, что болезни прогрессируют и развиваются преимущественно вдоль связей этой сети и, более того, было выявлено, что такое развитие различно для пациентов разного пола и различных рас. Было установлено также, что пациенты, страдающие заболеванием, имеющим большое число связей, умирают раньше, чем пациенты с заболеванием с небольшим числом связей. В этом исследовании была создана онлайн-запрашиваемая база данных из 18 различных наборов данных, полученных из истории болезней от более 31 миллиона пациентов. Связи (сообщества) заболеваний могут быть изучены в режиме онлайн или загружены в полном объеме.

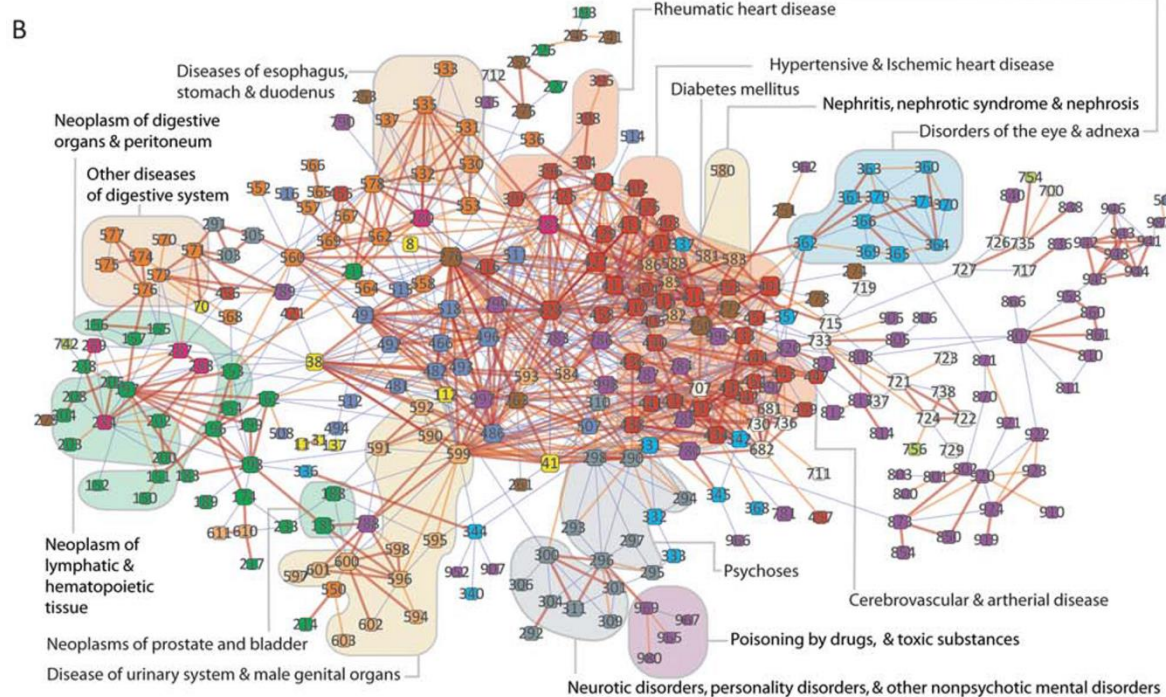
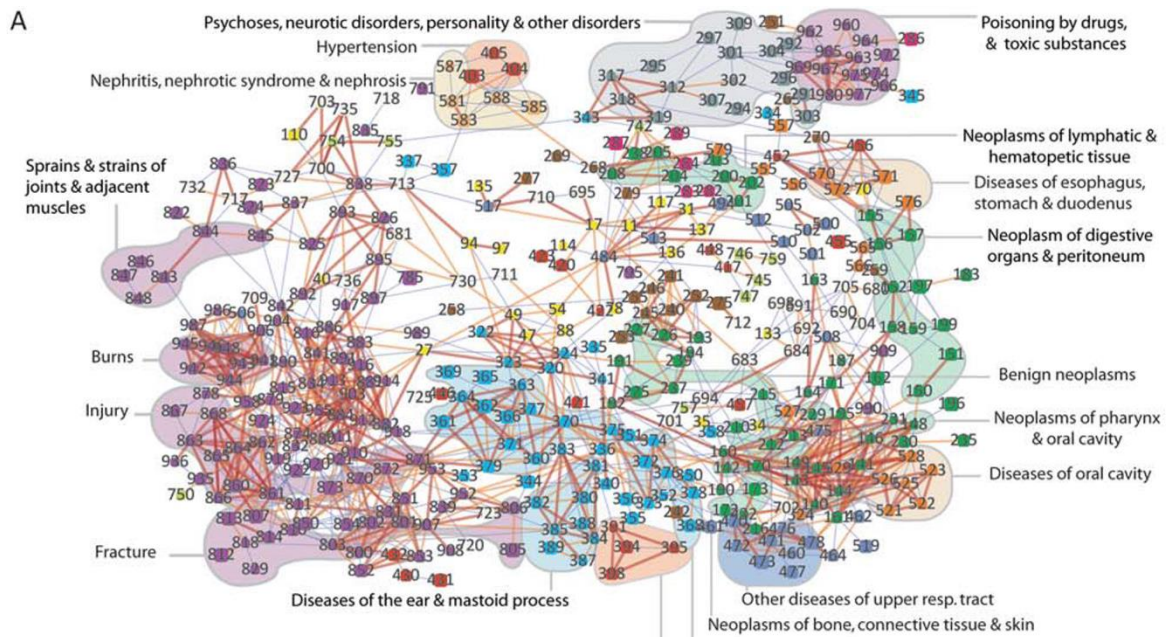
Обозначим  $RR_{ij} = \frac{C_{ij}N}{P_i P_j}$  – относительный риск заболеть двумя заболеваниями  $i$  и  $j$  для одного и того же пациента, где  $C_{ij}$  – число пациентов, страдающих обоими заболеваниями,  $N$  – полное число пациентов в популяции  $P_i$  и  $P_j$  число пациентов с заболеваниями соответственно  $i$  и  $j$ , то есть распространенность заболеваний  $i$  и  $j$ .

Корреляция Пирсона  $\Phi_{ij}$  как мера между двумя бинарными переменными:

$$\Phi_{ij} = \frac{C_{ij}N - P_i P_j}{\sqrt{P_i P_j (N - P_i)(N - P_j)}}$$

Эти две меры различия между сопутствующими заболеваниями (comorbidity) не являются абсолютно независимыми. Например, они обе растут с ростом числа пациентов, страдающих обоими заболеваниями.

При этом сеть, построенная на основе  $RR$  связей, переоценивает расстояния между редкими заболеваниями, в то время как  $\Phi$  недооценивает расстояния между парами редких заболеваний.



**Node Color**  
(ICD9 category) (a & b)

- 001-139 Infectious & Parasitic
- 140-239 Neoplasms
- 240-279 Endocrine, Nutritional, Metabolic & Immune
- 280-289 Blood & Blood-Forming Organs
- 290-319 Mental Disorders
- 320-389 Nervous System & Sense Organs
- 390-459 Circulatory System
- 460-519 Respiratory System
- 520-579 Digestive System
- 580-629 Genitourinary System
- 680-709 Skin & Subcutaneous Tissue
- 710-739 Musculoskeletal System & Connective Tissue
- 780-799 Symptoms, Signs, & ill-Defined Conditions

**RR Phenotypic Disease Network (a)**

- Node Size (Prevalence)**
- 1%
  - 0.1%
  - 0.01%
- Link weight**
- RR > 50
  - RR > 30
  - RR > 20

**Φ Phenotypic Disease Network (b)**

- Node Size (Prevalence)**
- 30%
  - 1%
  - 0.1%
- Link weight**
- φ > 0.2
  - φ > 0.1
  - φ > 0.06

**Рисунок 2.** (A) Фенотипическая Сеть Болезней (Phenotypic Disease Network – PDN), построенная по RR связям, образует своеобразное пространство фенотипов заболеваний (B). Фенотипическая Сеть Болезней (PDN), построенная по Φ связям.





В заключение, обсудим вопрос о связи смертности от заболевания  $i$  в зависимости от числа его связей с другими заболеваниями. Количественно число связей данного заболевания  $i$  вычисляется стандартными формулами теории сетей  $K_i^\Phi = \sum_j \Phi_{ij}$  и  $K_i^{RR} = \sum_j RR_{ij}$ .

Идеи и методы этого исследования были развиты для построения сети метастаз и создания модели развития ракового заболевания[3].

### Сети белковых взаимодействий при сахарном диабете второго типа

В статье [8] на основе многочисленных биологических баз данных построены сетевые структуры белок-белковых взаимодействий для клеток поджелудочной железы при сахарном диабете второго типа, проанализированы свойства состояний в норме и патологии и сделаны сравнения со свойствами случайных сетей Эрдеша-Реньи, сетей «тесного мира» Уоттса-Строгатца, а также сетей, построенных по конфигурационной модели, в которых сохраняется последовательность степеней узлов, следовательно, сохраняется и степенной закон сетей (scale free networks).

Распределение числа узлов по числу связей в сетях белок-белковых взаимодействий в клетках поджелудочной железы в норме и патологии является степенным. Такие сети со степенным законом распределения узлов по числу связей получили название безмасштабных сетей.

**Таблица 1.** Основные характеристики двух сетевых структур белок-белковых взаимодействий в норме ( $N_1$  и  $N_2$ ) и патологии ( $D_1$  и  $D_2$ ) при заболевании сахарным диабетом второго типа.  $N$  - общее число узлов (белков) в сети,  $N_c$  - общее число связей в сети,  $\langle k \rangle$  - среднее число связей узла сети,  $D$  - диаметр сети,  $\langle CC \rangle$  - среднее значение кластеризации сети.

Сеть	Число узлов	Число связей	Среднее число связей узла	Диаметр сети	Коэффициент кластеризации
$N_1$	2083	11017	10	17	0,35
$N_2$	1707	9888	11	12	0,33
$D_1$	656	3628	10	10	0,26
$D_2$	384	1882	10	9	0,46

Анализ свойств сетей белковых взаимодействий в норме и патологии показывает, что сети здоровых клеток поджелудочной железы являются более случайными, чем сети больных клеток и такое отсутствие достаточно высокой случайности сети, по-видимому и ведет к заболеванию.

Для подтверждения этой гипотезы целесообразно сравнить энтропии сетей белковых взаимодействий здоровых и больных клеток.

### Заключение

Описанные выше примеры далеко не исчерпывают основные направления приложения сетевых методов в медицине и биологии. Например, можно упомянуть статью, где описывается создание сетевой структуры «Лекарство-Мишень (Drug-Target)» [10]. Можно констатировать, что появившиеся в последние годы инструменты сетевой медицины на основе методов теории сложных сетей и современных баз данных взаимодействий биологических компонентов живой клетки на молекулярно-генетическом уровне формируют основу для систематического изучения не только сложных молекулярных механизмов конкретных заболеваний. Прогресс в этом направлении имеет важное значение и для выявления новых генетических причин болезней и ассоциированных с заболеваниями мутаций негенетического происхождения, а также создает предпосылки для более эффективного поиска лекарств и биомаркеров для сложных заболеваний.

В теории сложных сетей наиболее развитой динамической моделью эволюции сети является модель Бианкони-Барабаша (Bianconi-Barabasi)[1]. Эту модель часто еще называют «фитнес моделью», поскольку каждому узлу в ней присваивается на основании эмпирических данных фитнес-параметр, определяющий скорость роста числа связей у данного узла.

Как известно, процессы биологической эволюции относятся к феноменологии критических явлений и описываются математическими моделями на основе таких естественно-научными дисциплин как теория самоорганизованной критичности теория сложных сетей, теория катастроф, теория фракталов и т.п. Например, П.Бакпредложил модель пунктирного равновесия биологической эволюции на основе теории самоорганизованной критичности (СОК). Впервые теория старения на основе математического, физического и генетического подходов была предложена в 1959 году Лео Сцилардом [7]. Москалев А.А. предложил старение биологического организма описывать на основе понятия самоподобия (теория фракталов)[3]. В данной работе будет предложена модель старения (инволюции) на основе фундаментальных свойств сложных сетей, таких как робастность и каскадные процессы. Сразу же отметим, что безмасштабные сети обладают и свойством фрактальности [4]. Модель старения на основе безмасштабной сетевой модели фирм производящих модную одежду представлена в работах [1,5]. Однако, в этой модели отсутствует сама модель механизма старения. В настоящей работе мы представляем модель такого механизма в виде фундаментального свойства сетевых структур – каскадного процесса, впервые предложенного для объяснения масштабного отключения электроэнергии в сети электростанций (blackout) [1,6].

В реальных сетях активность одного узла может вызвать повреждение и отказ в работе связанных с ним узлов. То есть активность каждого узла зависит от активности его соседних узлов. Такие повреждения могут распространяться по сети, порождая, таким образом, каскад отказов в работе всех или большинства узлов в сети. Примером таких каскадов может служить выход из строя большого участка сети электростанций на севере США 14 августа 2003 года, что привело к прекращению подачи электричества около 45 миллионов жителей.

В работе [7] была представлена модель распространения повреждений в сети, которую можно кратко описать следующим образом. Рассмотрим сеть с произвольным законом распределения узлов по числу связей. Узел  $i$  может находиться в состоянии 0 (активное или здоровое состояние) или 1 (неактивное состояние) и характеризуется порогом  $\varphi_i$  выхода из строя. Вначале все узлы находятся в активном, здоровом состоянии 0. В момент времени  $t = 0$  один из узлов переключился из здорового состояния 0 в состояние 1. В следующий момент времени мы случайным образом выбираем узел и если этот узел находится в состоянии 0 исследуются все  $q_i$  его ближайших соседей и это узел переводится в состояние 1 если доля  $\varphi$  этих соседей находится в состоянии 1. В противном случае этот узел остается в состоянии 0. Если выбранный агент уже находится в состоянии 1, то его состояние не меняется. Иными словами, здоровый узел изменяет свое состояние, если доля  $\varphi_i$  его ближайших узлов находятся в нерабочем состоянии. В зависимости от локальной структура сети первоначальный выход из строя одного узла может никак не сказаться на здоровое состояние любого другого узла, но может привести к выходу из строя множества узлов. Такая модель описывает практически мгновенное наступление смерти в результате старения. Для реалистического постепенного старения каскадного процесса необходимо ввести параметр скорости старения, который может быть управляемой переменной.



## Литература

1. Barabasi A. Network Science. Cambridge: Cambridge University Press, 2016. 256 p.
2. Пер Бак. Как работает природа. М.: Либроком, 2008
3. Москалев А.А. Эволюционные представления о природе старения // Успехи геронтологии. 2010, т. 23 №1, с. 9-20
4. Song Ch., Havlin Sh., Makse H. Self-similarity of complex networks // Nature, January. 2005. Vol.23
5. Saavedra S., Reed-Tsochas, and Uzzi B. Assymmetric dissamly and robustness in declining networks // PNAS, October 28, vol 105 no43, 16486-16471
6. Watts. D.J. A simple model of global cascades on random networks // PNAS, 99: 5766, 2002.
7. Scilard L. On the nature of the aging process // PNAS, 1959, 45, 30-45
8. Яшин А.И. и др. Геронтология insilico: становление новой дисциплины // Успехи геронтологии, 2007, т.20, №1, с.7-19
9. Barabasi A.-L. Network Science. <http://barabasi.com/networksciencebook/>
10. Barabasi *et al.* Network medicine: a network based approach to human disease // Nature Review. Genetics, January 2011. Vol.12
11. Chen LL Blumm N, Barabasi A-L, Christakis NA and Deisboeck (2009). Cancer metastasis networks and the prediction of progression patterns // British Journal of Cancer, 2009. Vol 101, pp. 749–758.
12. Ghiassian, S. D., Menche, J., Barabási, A.-L. A DIseAseMOdule Detection (DIAMOnD) in Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome // PLOS Computational Biology DOI:10.1371. April 8, 2015
13. Goh *et al.* Humandiseasenetwork. PNAS. May 22, 2007, vol. 104, n. 21, 8687
14. Hidalgo C. A., Blumm, N., Barabasi A.L. & Christakis, N. A dynamic network approach for the study of human phenotypes // PLoS. Comput. Biol, 2009. Vol. 5, e1000353
15. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., Barabási, AL. Uncovering disease-disease relationships through the incomplete interactome // Science, 2015 • v. 347 ISSUE 6224, pp.841-848.
16. Rai, A., Pawar. K.A., and Jalan, S. Prognostic interaction patterns in diabetes mellitus II: A random-matrix-theory relation // Physical Review E 92, 2015. 022806
17. Rzhetsky, A., Wajngurt, D., Park, N & Zheng, T. Probing genetic overlap among complex human phenotypes // PNAS, 2007. **104**, 11694-11699.
18. Yildirim M., Goh K., Gusick M., Barabasi AL., Vidal M. Drug-target network // Nature Biotechnology. Oktober 2007. 25, N 10.
19. Zhou X. Z., Mencht J., Barabasi A.L., Sharma A.. Human symptoms-disease network // Nature Communications. 2014, 5-4212.DOI: 10.1038.