

# Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

[bartek@mimuw.edu.pl](mailto:bartek@mimuw.edu.pl)

<http://www.mimuw.edu.pl/~bartek>

**Wykład 4. - Narzędzia do wizualizacji  
Cytoscape I Circos**

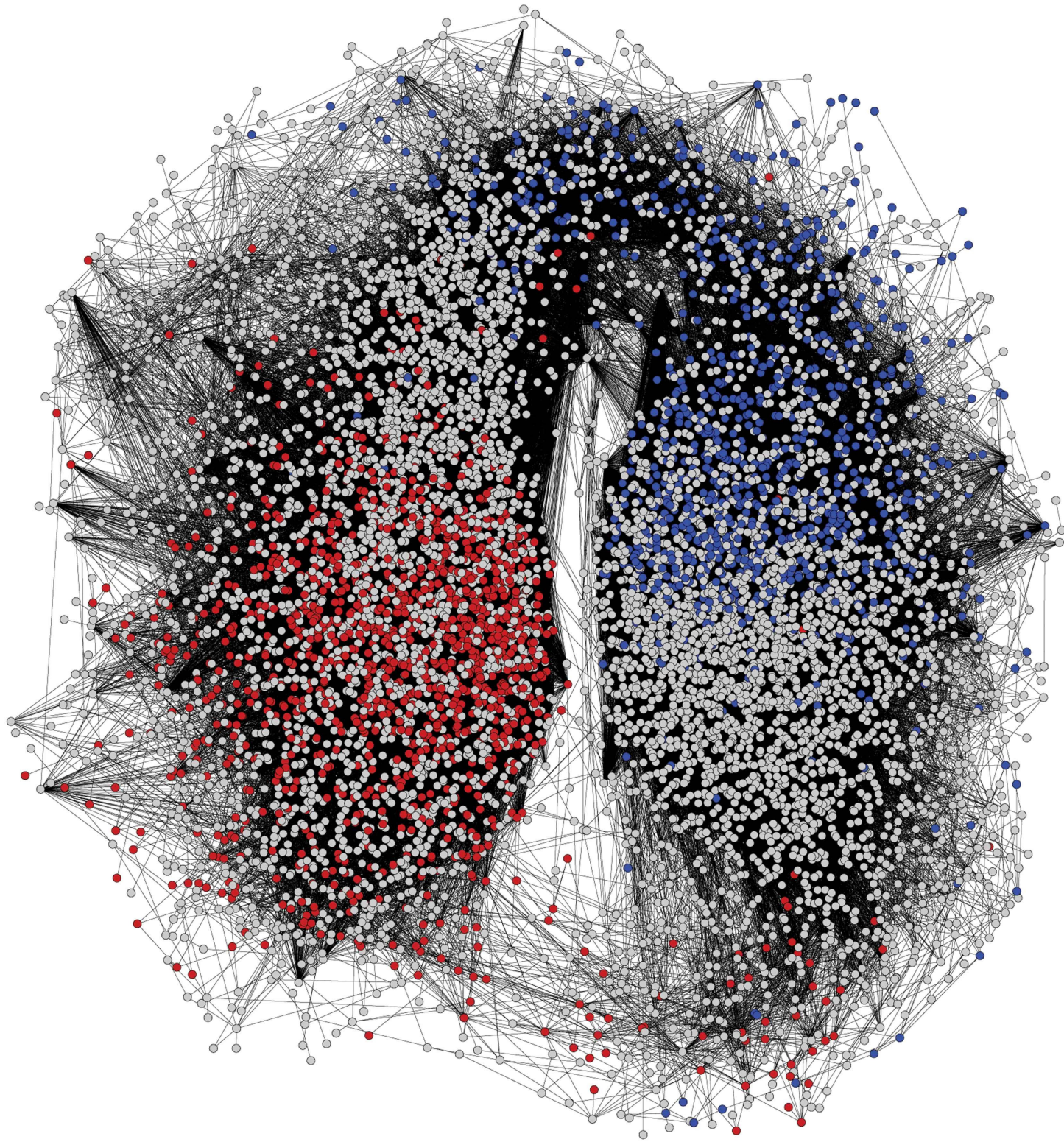
26. X. 2015

# Tematy na dziś

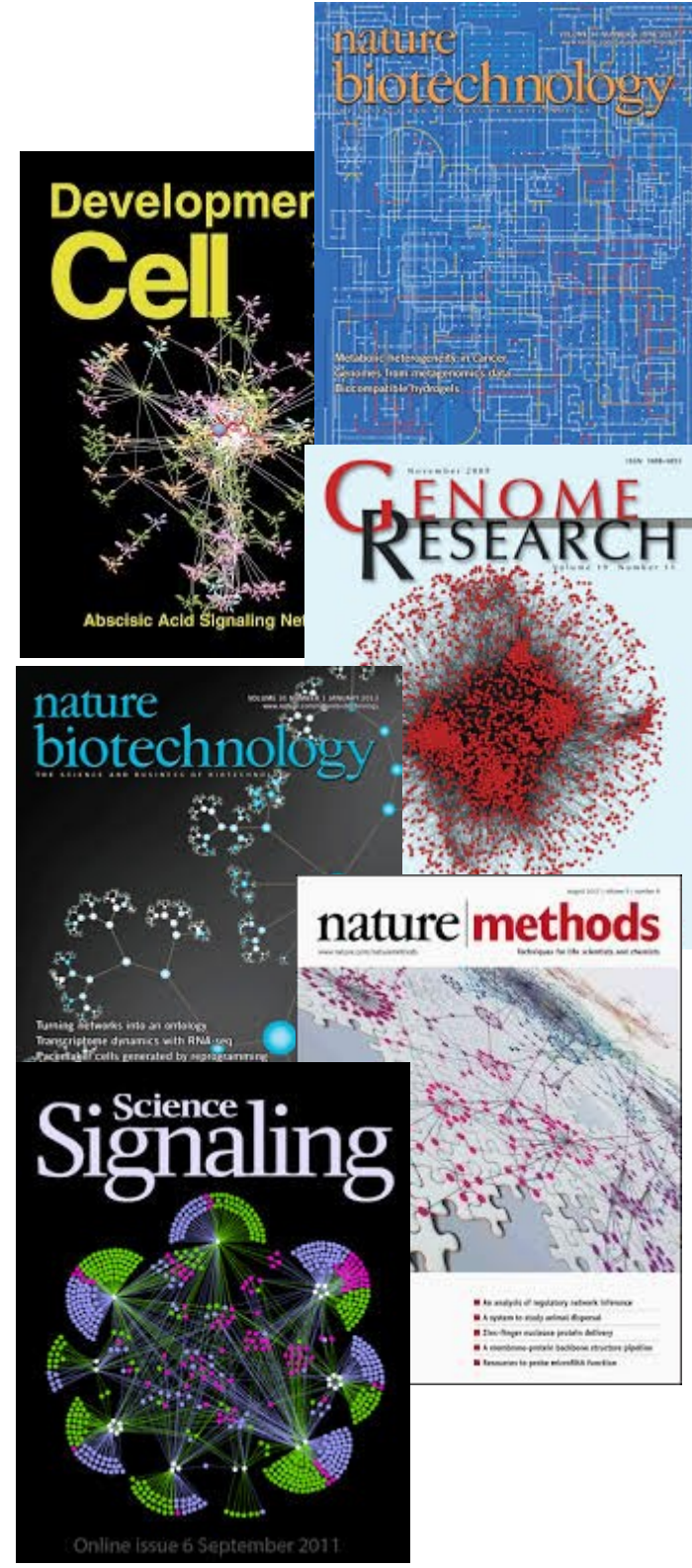
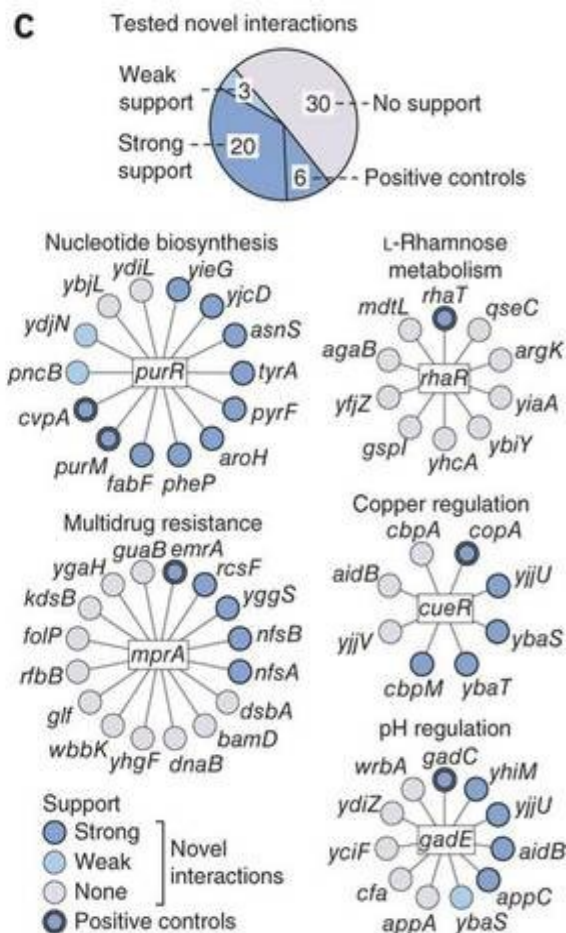
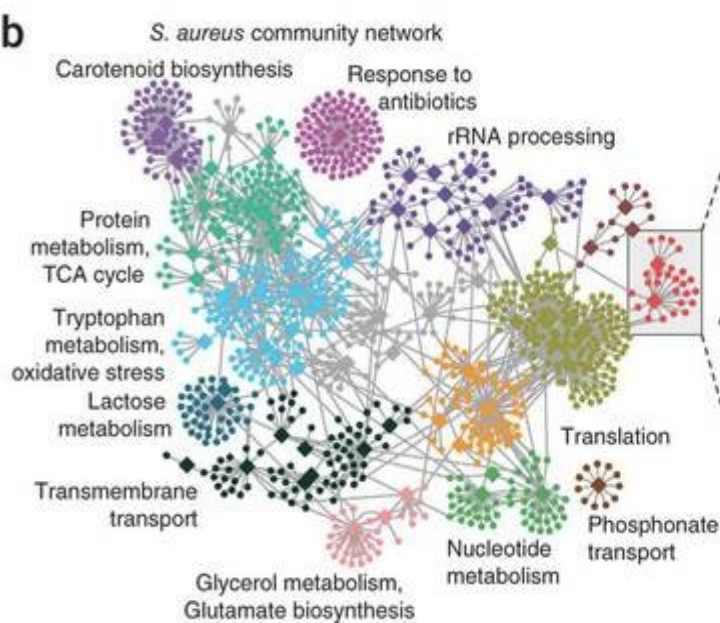
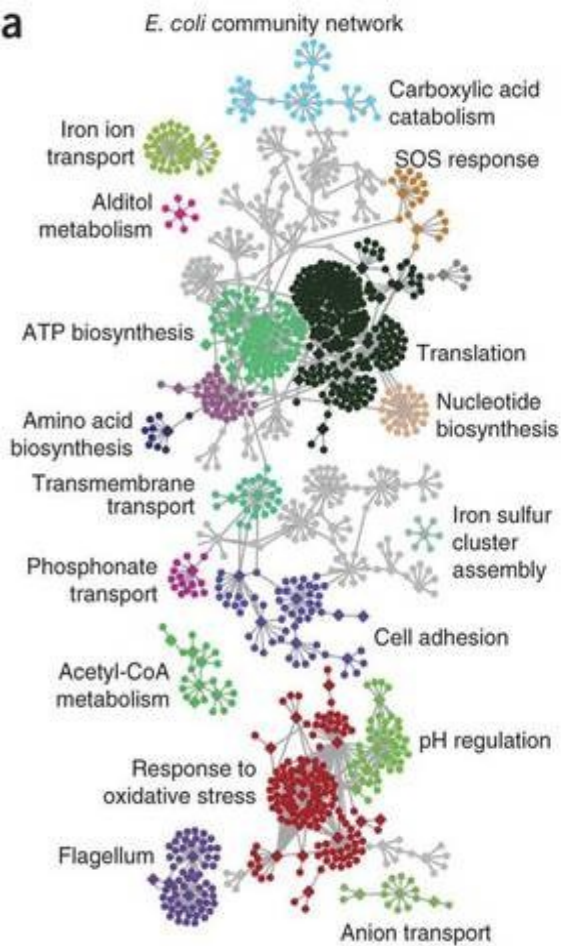
- O sieciach w bioinformatyce
- Cytoscape – trochę historii (Trey Ideker)
- Pomysł na Cytoscape
- Plugin'y w wersji 2.0
- Appstore w wersji 3.0
- Cytoscape.js (dawniej cytoscape-web)
- Circos – zupełnie inny sposób działania
- Circos online – przeglądarki są coraz bardziej popularne...

# Sieci w bioinformatyce

- Sieci interakcji genów
- Sieci interakcji białek
- Sieci regulacji genów
- Szlaki metaboliczne
- Sieci interakcji chromosomowych
- Acykliczne grafy skierowane (DAG) ontologii
- Sieci cytowań artykułów
- Sieci interakcji asocjacyjnych (z GWAS)
- Sieci targetów farmakologicznych
- Sieci ko-ekspresji, ko-regulacji, ko-ewolucji...







# Dlaczego warto badać sieci?

- Ogromna ilość informacji o interakcjach ( $O(n^2)$ )
- Trudna do objęcia inaczej niż w komputerze
- Pewne własności pojawiają się w wielu różnych rodzajach sieci
- Np. częste występowanie wierzchołków o niespodziewanie wysokim stopniu (tzw. Hub'y)
- Bez zaawansowanych narzędzi do wizualizacji nie jest łatwe analizowanie dużych sieci





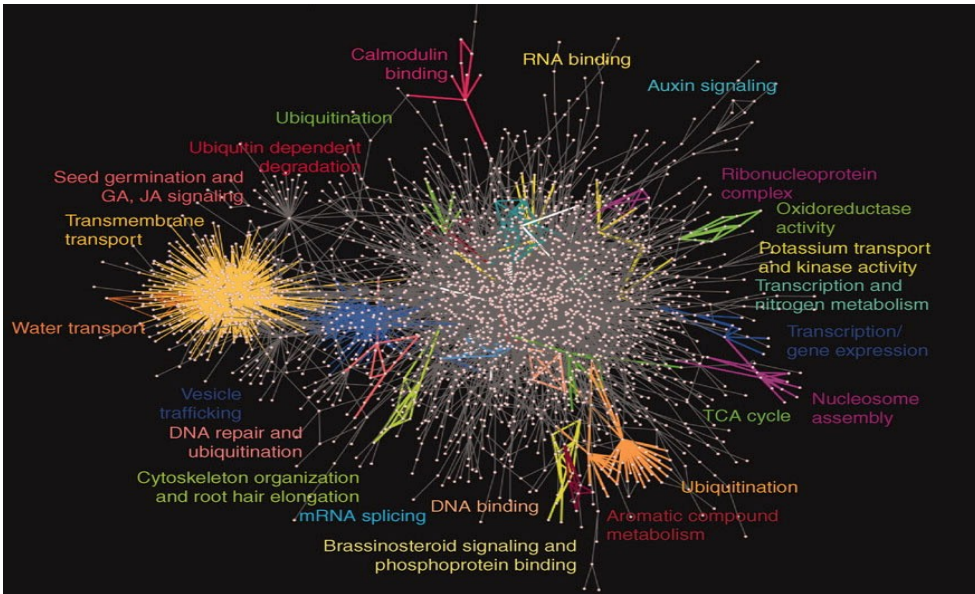








# Sieci małego świata





# Cytoscape

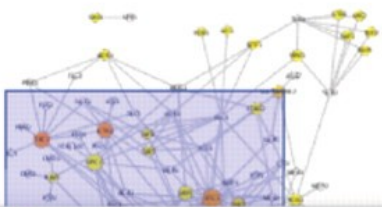
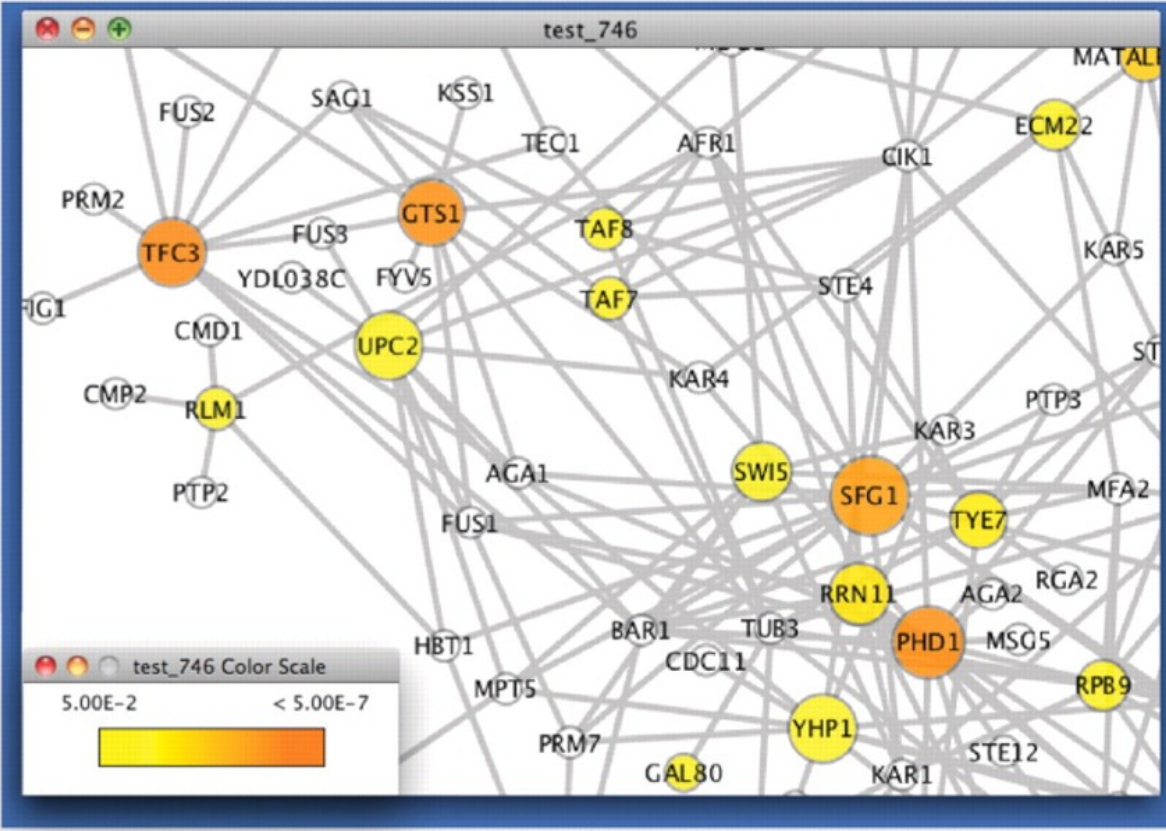
- Projekt rozpoczęty przez Trey'a Ideker'a w Institute for Systems Biology w 2002 r
- Wersja 1.0 w 2003
- Interfejs użytkownika w Javie
- Używanie bibliotek OpenSource do wizualizacji grafów
- Podstawowe operacje na grafach
- Wsparcie dla formatów danych używanych w biologii



Control Panel

Network VizMapper™

Network	Nodes	Edges
test_746	90(0)	186(0)



PiNGO output

test\_746

/Users/maere/Documents/go/gene\_association.sgd, /Users/maere/Documents/go/gene\_ontology.obo close

GO ID	GO Description	Gene name	Gene alias	Gene description	Neighbor freq	Data freq	P value	Existing annotations
<input type="checkbox"/>	746 conjugation	TFC3	TFC3	Largest of six subunits of the RNA pol...	68.7 11/16	1.8 119/6502	2.2093E-14	protein localization to chromatin(71168) reg...
<input type="checkbox"/>	746 conjugation	PHD1	PHD1	"Transcriptional activator that enhance...	12.1 13/107	1.8 119/6502	1.0709E-6	translational elongation(6414) transcription(...
<input type="checkbox"/>	746 conjugation	GTS1	GTS1	"Arf3p GTPase Activating Protein (GAP...	17.2 10/58	1.8 119/6502	1.1588E-6	zinc ion binding(8270) chronological cell agl...
<input type="checkbox"/>	746 conjugation	SFG1	SFG1	"Nuclear protein, putative transcrip...	7.7 17/218	1.8 119/6502	4.5208E-6	transcription(6350) positive regulation of tra...
<input type="checkbox"/>	746 conjugation	MATALPHA2	MATALPHA2	"Homeobox-domain protein that, with...	36.3 4/11	1.8 119/6502	2.5590E-4	donor selection(7535) regulation of transcrip...
<input type="checkbox"/>	746 conjugation	RRN11	RRN11	"Component of the core factor (CF) rD...	7.4 8/108	1.8 119/6502	3.7409E-3	transcription of nuclear rRNA large RNA poly...
<input type="checkbox"/>	746 conjugation	YHP1	YHP1	"One of two homeobox transcriptional...	4.2 11/261	1.8 119/6502	1.9663E-2	nuclear chromatin(790) specific transcrip...
<input type="checkbox"/>	746 conjugation	RPB9	RPB9	RNA polymerase II subunit B12.6; con...	4.3 10/230	1.8 119/6502	2.0798E-2	DNA-directed RNA polymerase activity(3899...
<input type="checkbox"/>	746 conjugation	MTF1	MTF1	"Mitochondrial RNA polymerase specif...	7.5 4/53	1.8 119/6502	2.9245E-2	DNA binding(3677) transcription initiation fr...
<input type="checkbox"/>	746 conjugation	RLM1	RLM1	"MADS-box transcription factor, comp...	5.8 5/86	1.8 119/6502	3.2727E-2	DNA bending activity(8301) response to acid...
<input type="checkbox"/>	746 conjugation	TOA2	TOA2	"TFIIA small subunit; involved in transc...	5.0 6/120	1.8 119/6502	3.4603E-2	regulation of transcription(45449) transcrip...
<input type="checkbox"/>	746 conjugation	WTM2	WTM2	"Transcriptional modulator involved in...	6.7 4/59	1.8 119/6502	3.4603E-2	transcription activator activity(16563) transcr...
<input type="checkbox"/>	746 conjugation	YOX1	YOX1	"Homeodomain-containing transcrip...	3.4 12/346	1.8 119/6502	3.5526E-2	nuclear chromatin(790) specific transcrip...
<input type="checkbox"/>	746 conjugation	TAF8	TAF8	"TFIID subunit (65 kDa), involved in R...	4.9 5/101	1.8 119/6502	4.2610E-2	interphase(51325) regulation of transcrip...
<input type="checkbox"/>	746 conjugation	TAF7	TAF7	"TFIID subunit (67 kDa), involved in R...	4.9 5/101	1.8 119/6502	4.2610E-2	SAGA complex(124) regulation of transcrip...

Select All Unselect All

Select nodes

# Wersja 2.0

- Wydana już w roku 2004
- Rozwijana przez grupę Trey'a Idekera (UCSD) ze wsparciem innych ośrodków badawczych (San Francisco, Toronto, etc.)
- Wsparcie dla “wtyczek” (plug-ins)
- Ogromny sukces (tysiące cytowań, dziesiątki wtyczek)
- Problemy z kompatybilnością (wersje javy, wersje cytoscape, częste odwołania do zewnętrznych narzędzi)

# Cytoscape-web

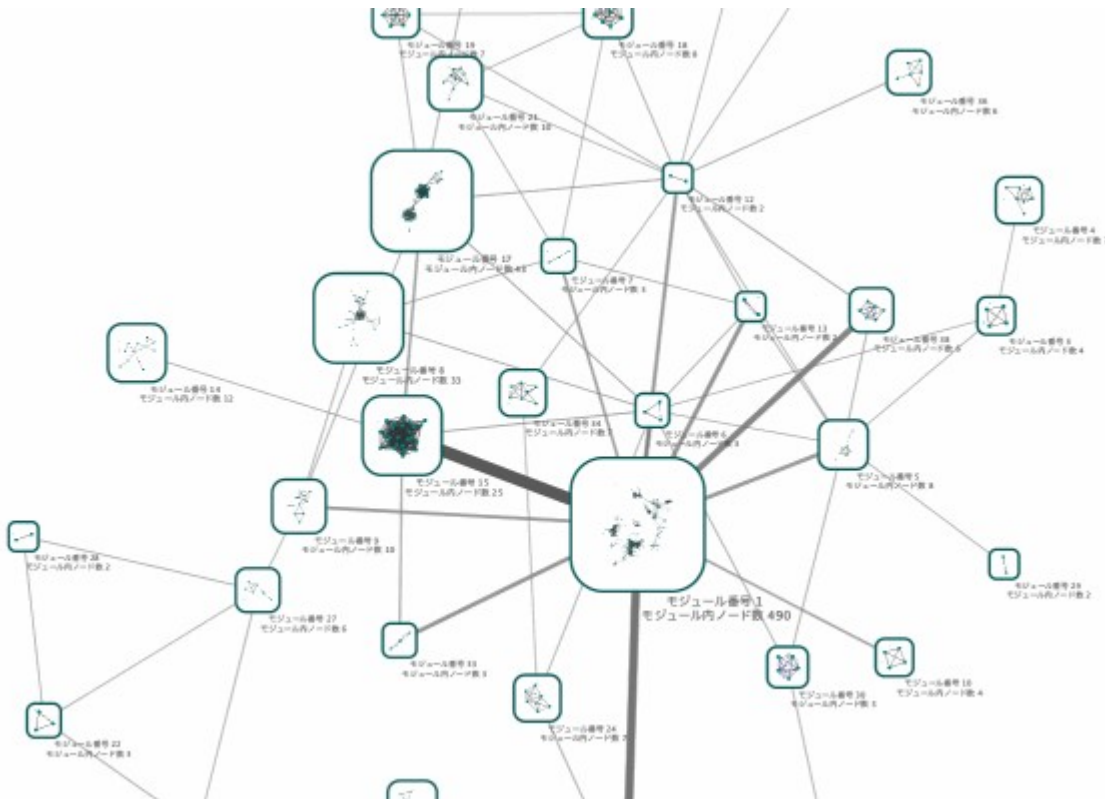
- Potrzebne wersje cytoscape dostępne online
- Dla osób, które nie chcą instalować cytoscape i dla badaczy, którzy chcieliby udostępniać swoje wyniki “publiczności” na stronie www
- Cytoscape-web 2007-2013

The screenshot displays the Cytoscape-web interface. At the top, there is a search bar containing 'atg17 atg16' and buttons for 'Search', 'Clear', and 'Show Advanced Fields'. A 'MIQL syntax reference' link is also present. Below the search bar is a navigation menu with tabs for 'Home', 'Search', 'Interactions (189)', 'Browse', 'Lists', 'Interaction Details', 'Molecule View', and 'Graph'. The 'Graph' tab is selected and highlighted with a red box labeled 'A'. The main area shows a 'Network visualisation' of a star graph with a central node and many peripheral nodes. To the right of the graph is a control panel with two sections: 'Open in Cytoscape' (containing a Cytoscape icon and text about opening the search in a WebStart version of Cytoscape 2.6.3) and 'CytoscapeWeb Controls' (containing layout options like 'force directed', 'radial', and 'circle', and a 'Merge edges' toggle set to 'on'). A red box labeled 'B' highlights the 'Open in Cytoscape' section, and a red box labeled 'C' highlights the 'CytoscapeWeb Controls' section.



# Cytoscape.js

- w pełni HTML5, od 2011
- Pozwala na “wbudowywanie” sieci do aplikacji Java sctiptowych



# Cytoscape 3.0

- Aktualna wersja Cytoscape
- W pełni modularna budowa, lepiej dopracowane API
- Odejście od numerów wersji (problem z kompatybilnością np. 2.7 vs. 2.8 w Cytoscape 2)
- App-store zamiast wtyczek
- Cytoscape consortium zyskało status własnej organizacji not-for-profit
- Interfejs użytkownika właściwie bez zmian

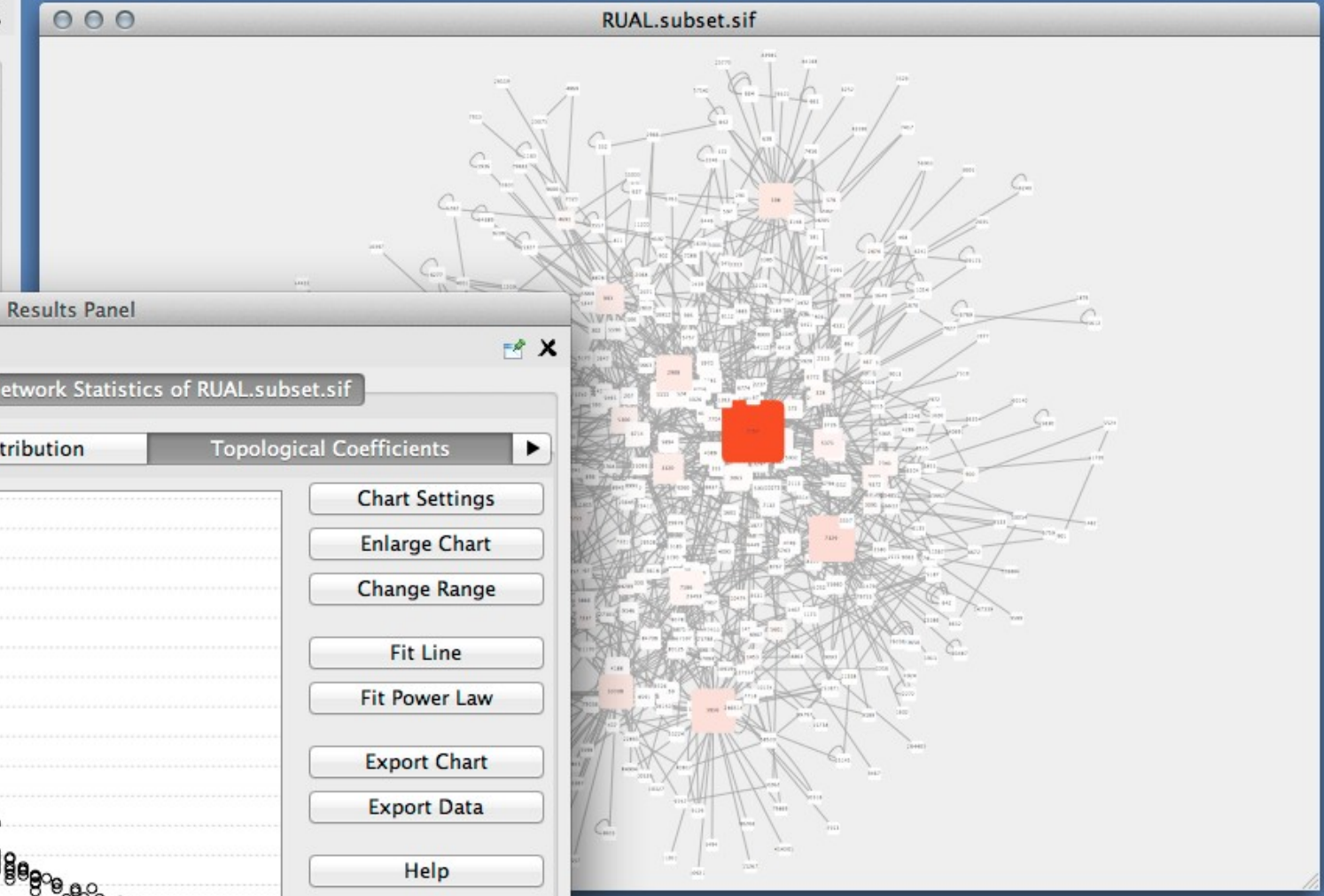
Control Panel

Network VizMapper Filters

Current Visual Style

Minimal

Defaults (Click to edit)



Results Panel

Node Details Network Statistics of RUAL.subset.sif

Avg. Clustering Coefficient Distribution

Topological Coefficients

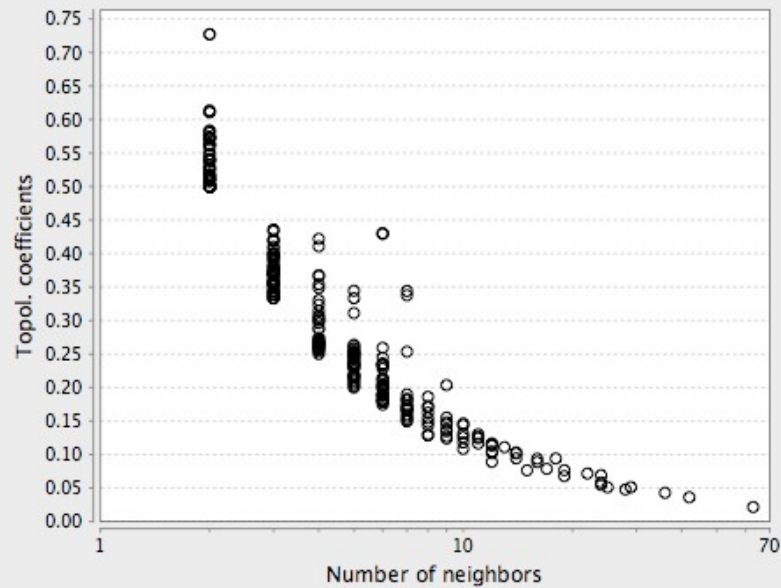


Chart Settings

Enlarge Chart

Change Range

Fit Line

Fit Power Law

Export Chart

Export Data

Help

Save Statistics

Visualize Parameters

Close Tab

RUAL.subset.sif

Node Table Edge Table Network Table

# Przykłady użycia

- Nexo – automatycznie budowane ontologie
- J. Dutkowski, 2012
- Technologia dostępna na licencji BSD, komercjalizowana w firmie data4life

NeXO About User Manual Developer Manual Policy Download Team

## What is NeXO?

**NeXO** is a gene ontology inferred directly from large-scale molecular networks. A gene ontology provides structured knowledge about the cellular components, processes, and functions encoded by genes. While most ontologies – including the highly successful [Gene Ontology Database \(GO\)](#) – are constructed through manual expert curation, NeXO is a data-driven gene ontology inferred directly from 'omics data.

**NeXO (The Network Extracted Ontology)** uses a principled computational approach which integrates evidence from hundreds of thousands of individual gene and protein interactions to construct a complete hierarchy of cellular components and processes. This data-derived ontology aligns with known biological machinery in the GO Database and also uncovers many new structures.

**The NeXO approach is described in the following publication:**

[A gene ontology inferred from molecular networks. Janusz Dutkowski, Michael Kramer, Michal A Surma, Rama Balakrishnan, J Michael Cherry, Nevan J Krogan & Trey Ideker. Nature Biotechnology 31, 38–45 \(2013\).](#)

**The Gene Ontology data sets used in this web application is maintained by The Gene Ontology Consortium:**

[Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. May 2000;25\(1\):25-9.](#)

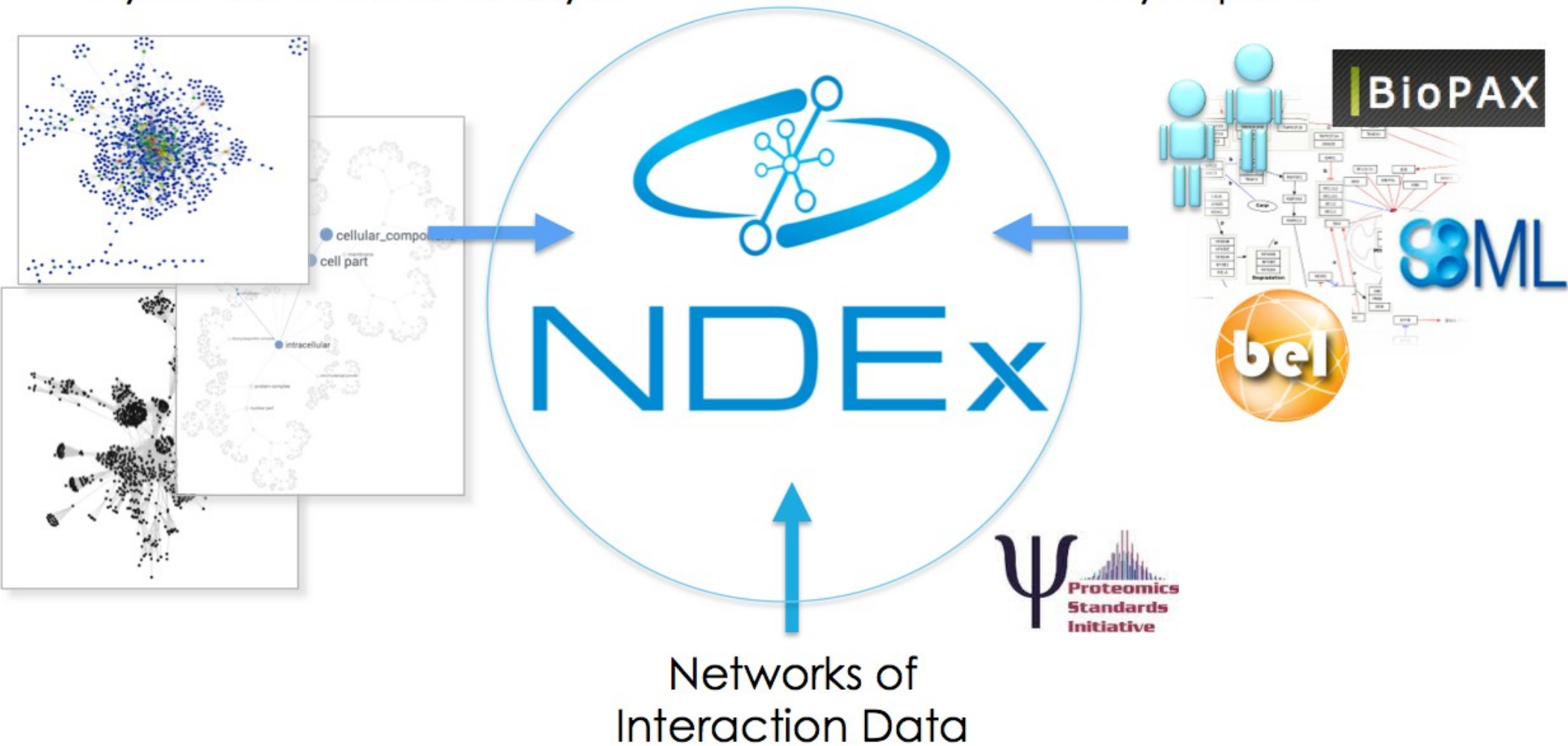




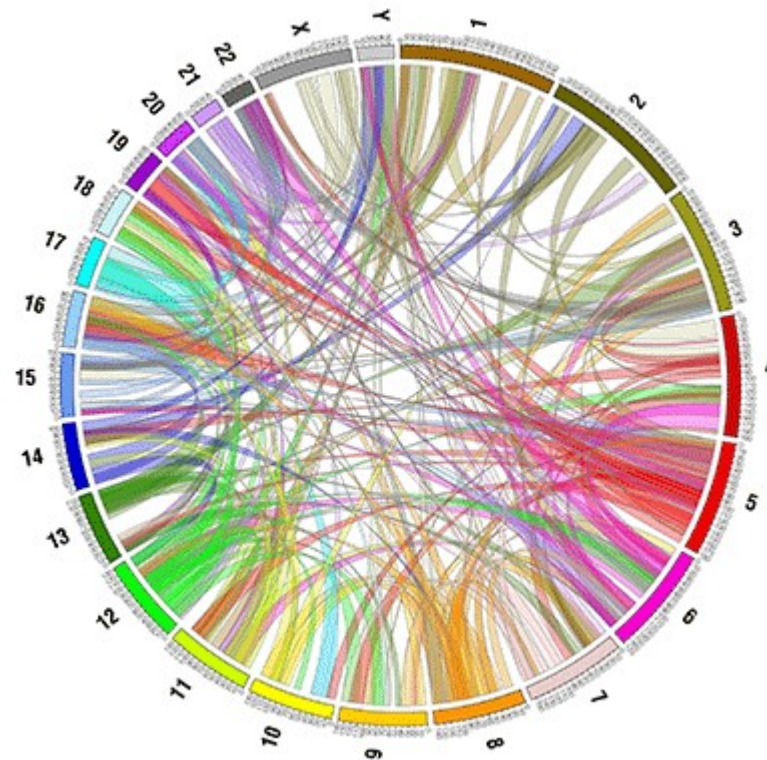
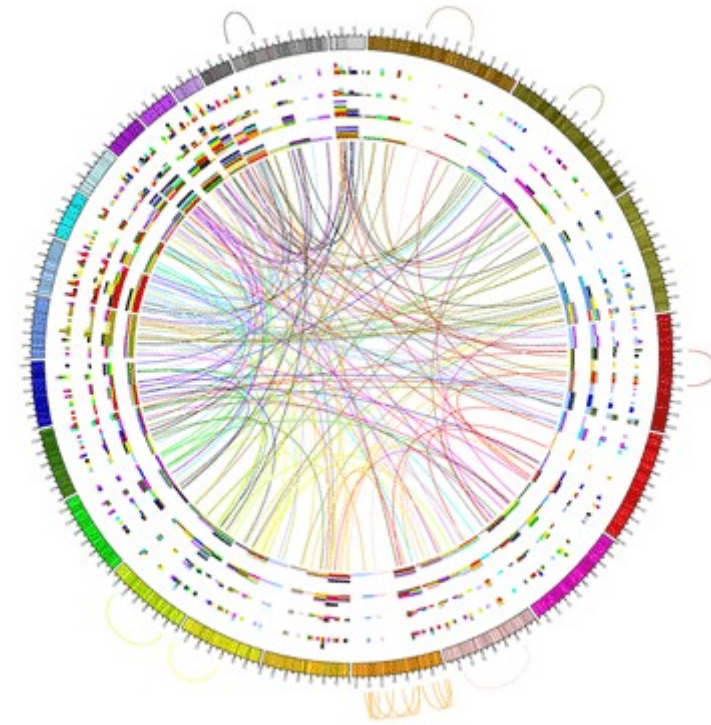
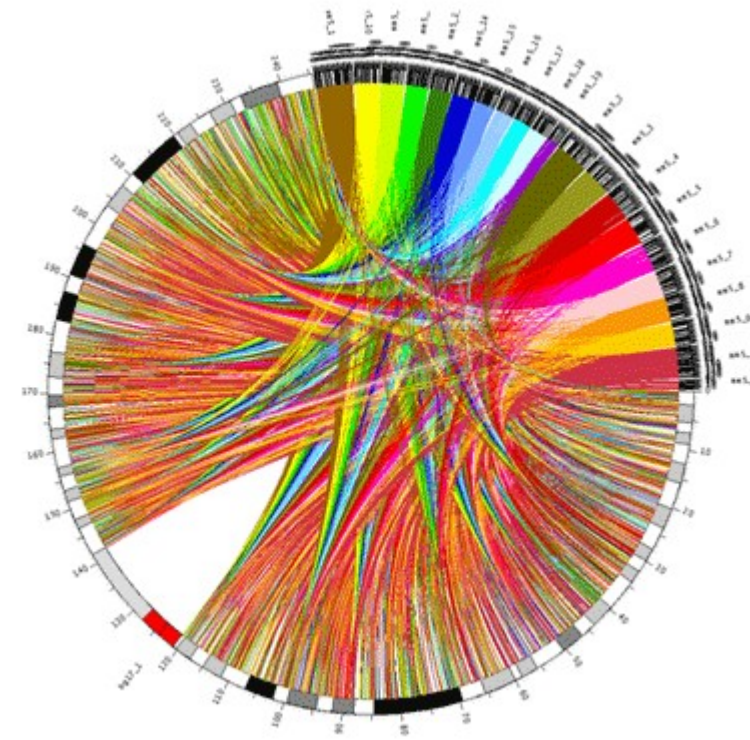
# NDEx – Network data exchange

Networks from  
Systematic Data Analysis

Networks Curated  
by Experts



# Circos



# Zupełnie inny pomysł - Circos

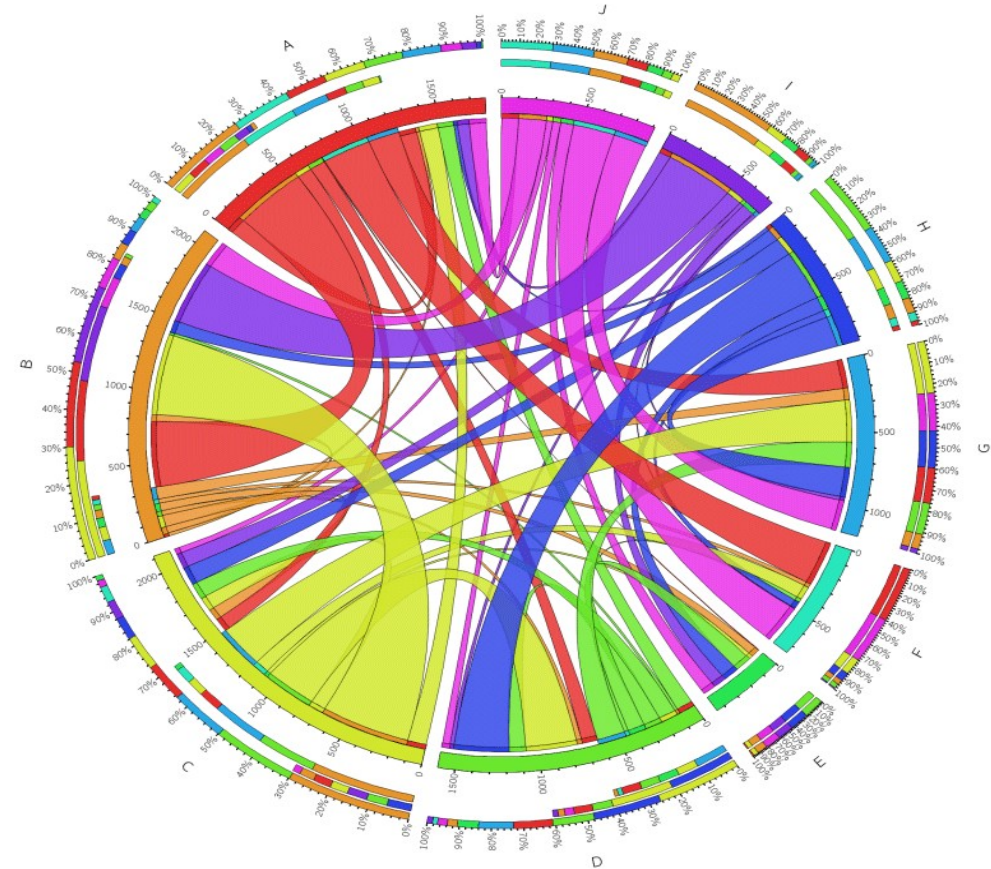
- Martin Krzywinski – od 2004 roku
- Jeden developer
- Zestaw skryptów zamiast GUI
- Pliki konfiguracyjne
- Inkrementalny postęp, próby zachowania zgodności wstecz
- Duże wysiłki w promocję tego pomysłu
- Zastosowania także do innych typów danych



# Circos online

# Medium tab-delimited table with mixed row and column labels

data	A	B	C	D	E	F	G
A	105	450	92	96	5	301	195
B	20	46	78	33	53	28	83
C	118	553	94	317	25	89	287
D	100	18	108	104	105	25	173
H	23	83	123	342	98	48	205
I	73	428	103	25	82	5	23
J	105	173	38	49	81	258	207





# Biblioteki do analiz sieci

- Graphviz – stara uniksowa biblioteka (wizualizacje podobne do cytoscape, interfejs bardzo podobny do Circos)
- Networkx – biblioteka do “przetwarzania” sieci, wyjście najczęściej do innych formatów (graphviz, cytoscape)

# Interactive tree of life

File Edit View History Bookmarks Tools Help

itol: Interactive Tre... x

itol.embl.de

HOME TREE OF LIFE OTHER TREES SHARED PROJECTS DATA UPLOAD SAVED VIEWS HELP ABOUT & CONTACT Login

## ITOL INTERACTIVE TREE OF LIFE

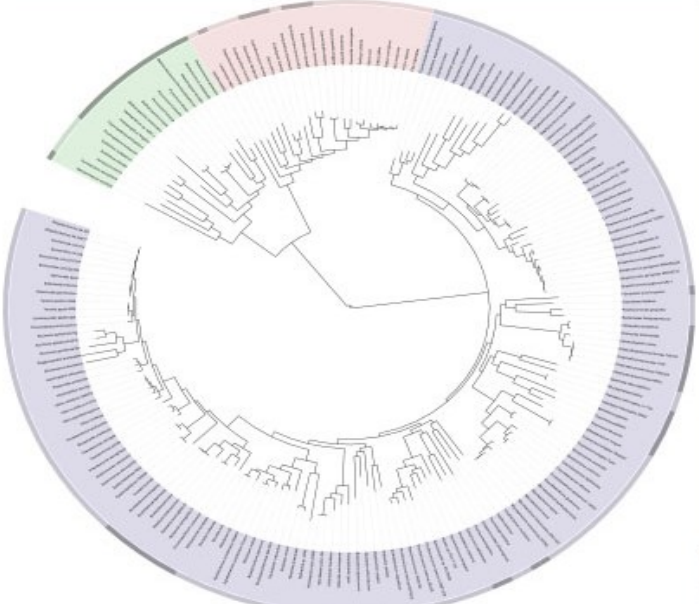
### Welcome to iTOL!

**Interactive Tree Of Life** is an online tool for the display and manipulation of phylogenetic trees. It provides most of the features available in other tree viewers, and offers a novel circular tree layout, which makes it easy to visualize mid-sized trees (up to several thousand leaves). Trees can be exported to several graphical formats, both bitmap and vector based. [more...](#)

#### News

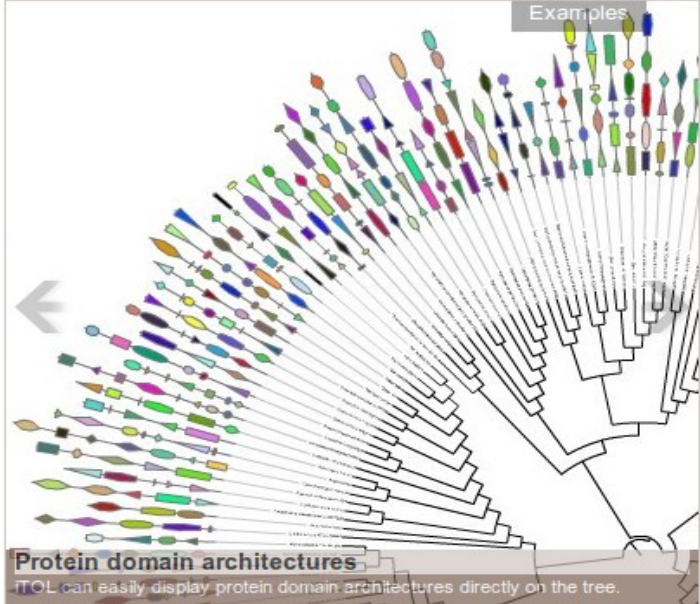
- Version 2.1 introduces a new dataset type, **circles**. Recent additions also include the ability to assign dataset values with internal nodes. These will be displayed only when associated clades are collapsed. Check the [Help pages](#) for details.
- Second iTOL article was published in 2011 NAR Web server issue ([abstract](#), [full text PDF](#)).

#### The Tree Of Life



#### Various iTOL generated tree images

Examples



**Protein domain architectures**  
iTOL can easily display protein domain architectures directly on the tree.

#### iTOL account login

Login:  Password:

Remain logged-in?

[No account?](#)

#### Firefox toolbar

If you are using Mozilla Firefox to access iTOL, try our [Firefox toolbar](#).

#### Recent changes

**Version 3.0**

- ↳ Complete rewrite of the backend tree storage system
- ↳ **Additional tree** can be displayed above the heatmap datasets
- ↳ Support for **display of custom scales** in several dataset types
- ↳ Various minor updates and bug fixes

**Version 2.2.2**

- ↳ Collapsed clades can be exported with same size and font as regular leaves

**Version 2.2.1**

- ↳ Colored strip and gradient datasets support values assigned to internal tree nodes. These values will only be displayed when associated clades are collapsed.
- ↳ Collapsed clades can be displayed with simplified triangles, taking less space

**Version 2.2**

- ↳ multiple binary datasets will be automatically spread to prevent overlap
- ↳ If labels are present in a color strip dataset, a legend containing the labels will be added to exported trees

**Version 2.1.1**

- ↳ branches can be marked with **custom labels**
- ↳ Dataset legends are optional in exported trees

**Version 2.1**

- ↳ New dataset type: **circles**
- ↳ Several datasets (binary, bar chart, multi-value bar chart, protein domains) support display of values assigned with internal nodes of the tree. These values will only be displayed when associated clades are collapsed.

**Version 2.0.1**

- ↳ Pie chart position on each branch can be defined: [example](#)

[Full version history](#)



# Biobyte – firma stojąca za iTOL

File Edit View History Bookmarks Tools Help tools, services and consulting - Mozilla Firefox

biobyte solutions G... x +

www.biobyte.de Search

biobyte SOLUTIONS

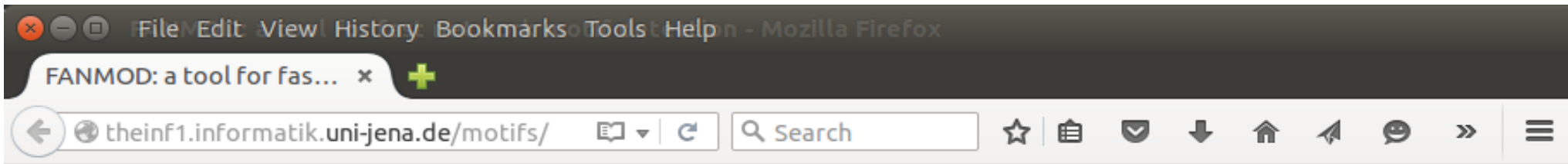
Home About Us Services Tools Contact

BIOINFORMATICS SERVICES / TOOLS / CONSULTING

## TOOLS & SERVICES

ABOUT US >

# Motywy w sieciach



## FANMOD a tool for fast network motif detection

The screenshot shows the FANMOD software interface. The main window is titled 'FANMOD' and contains several sections:

- SETUP:** Includes 'Algorithm options' (Subgraph size: 4, Samples to estimate number of subgraphs: 10000), 'Random networks' (Number of nodes: 429, Exchanges per edge: 5, Exchange attempts: 3), and 'ALGORITHM' (Start button, Progress bar).
- RESULTS:** Displays network statistics: 'Network name: C:\Exkursions und Erdbeerkugeln\erdbeerkugeln\29962\29962\coll1\_1inter\_st.txt', 'Network type: Directed', 'Number of nodes: 429', 'Number of edges: 529', 'Number of single edges: 539', 'Number of isolated edges: 0', 'Number of edge colors: 3'. It also has 'HTML Export' and 'Load File' buttons.
- HTML Export options:** A dialog box for saving results as HTML, including 'File Options' (HTML-Output-File: C:\projekte\coll1\_1inter\_st.html), 'Motifs per file: 20', and 'Filters' (E-score bigger than: 2, p-value less than: 0.05, Frequency bigger than: 0.01, Most frequent or least: 5 times, Ignore motifs that have between: 1 and 1 dangling edges).
- HTML Output:** A browser window showing 'Size-4 Network Motifs in Network 'coll1\_1inter\_st.txt' - Page 1 - Generated With FANMOD - Open'. It contains a table of motifs with their IDs, adjacency matrices, and statistical data.

ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-score	p-value
205		0.007152%	2.352e-006%	5.2569e-007	135	0
206		0.004768%	2.3727e-006%	5.3026e-007	89.67	0
2188		0.002934%	1.6029e-006%	2.7304e-007	62.739	0
205		0.003570%	3.5635e-006%	6.4929e-007	54.961	0

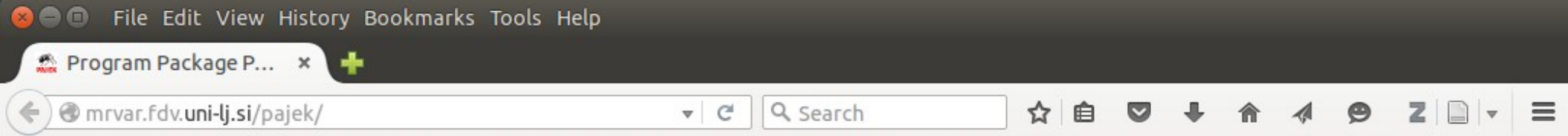
## Introduction

From this page you can download a tool to detect network motifs as described in

- S. Wernicke and F. Rasche.  
**FANMOD: a tool for fast network motif detection.**




# Pajek – narzędzie dla windows (free for noncommercial use)

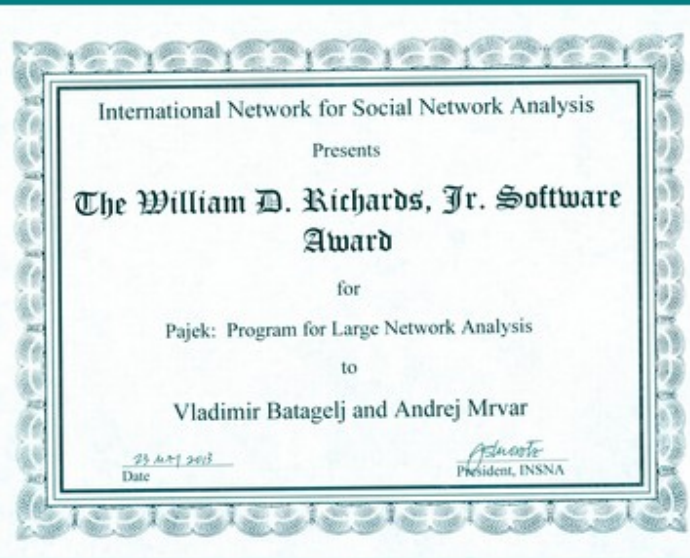


## Pajek / Pajek-XXL versions 3.\*\* and 4.\*\*

### Download - August 25, 2015

	<b>32 bit</b>	<b>64 bit</b>
<b>August 25, 2015</b>	<b><u>4.05</u></b>	<b><u>4.05</u></b>
June 9, 2015	<u>4.04</u>	<u>4.04</u>
September 25, 2011	<u>2.05</u>	<u>2.05</u>

[Pajek mailing list](#)




### First steps




**Pajek  
and  
Pajek-XXL**  
Programs for Analysis

STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES 32

**Exploratory  
Social Network  
Analysis with Pajek**  
REVISED AND EXPANDED



Exploratory  
Social Network  
Analysis with Pajek

Pajekを活用した  
社会ネットワーク  
分析



**PAJEK**  
蜘蛛:  
社会网络分析技术

Exploratory Social Network Analysis with Pajek (Second Edition)

CAMBRIDGE