Archive of SID

ISSN: 2476-6917; Modares Journal of Biotechnology. 2019;10(2):329-334

PRAF Framework for Global Protein-Protein Interaction Network Alignment

ARTICLE INFO

Article Type
Original Research

Authors
Ghorbanali Z.[1] *MSc,*
Zare-Mirakabad F.*[1] *PhD*

How to cite this article
Ghorbanali Z, Zare-Mirakabad F. PRAF Framework for Global Protein-Protein Interaction Network Alignment. Modares Journal of Biotechnology. 2019;10(2): 329-334.

ABSTRACT

A biological network represents the interaction between a set of macromolecules to drive a particular biological process. In a biological environment, abnormalities happen not only in one molecule but also through a biological network. One of the most effective methods to detect anomaly is the comparison between healthy and diseased networks. In this regard, biological network alignment is one of the most efficient ways to find the difference between healthy and diseased cells. This problem, protein-protein interaction network alignment, has been raised in two main types: Local network alignment and Global network alignment. According to the NP-completeness of this problem, different non-deterministic approaches have been proposed to tackle the Global network alignment problem. Recently, NetAl has been introduced as a common algorithm to align two networks. Although this algorithm can align two networks at the appropriate time, it does not consider biological features. In this study, we present a new framework called PRAF to improve the results of network alignment algorithms such as NetAl by considering some biological features like gene ontology (GO).

Keywords Bioinformatics; Protein-protein interaction networks; Alignment; Gene ontology; Clustering

[1]Computer Science Department, Mathematics & Computer Science Faculty, Amirkabir University, Tehran, Iran

Correspondence
Address: Computer Science Department, Mathematics & Computer Science Faculty, Amirkabir University, Tehran, Iran
Phone: +98 (21) 64545674
Fax: -
f.zare@aut.ac.ir

Article History
Received: February 4, 2018
Accepted: March 3, 2018
ePublished: June 20, 2019

CITATION LINKS

[1] Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks [2] Yeast two-hybrid: so many interactions, (in) so little time [3] Mass spectrometry-based proteomics [4] The protein threading problem with sequence amino acid interaction preferences is NP-complete [5] GASOLINE: a Greedy and Stochastic algorithm for Optimal Local multiple alignment of Interaction Networks [6] Global alignment of multiple protein interaction networks with application to functional orthology detection [7] BinAligner: A heuristic method to align biological networks [8] NETAL: A new graph-based method for global alignment of protein-protein interaction networks [9] Clustermaker: A multi-algorithm clustering plugin for Cytoscape [10] Detecting overlapping protein complexes in protein-protein interaction networks [11] CytoCluster: A cytoscape plugin for cluster analysis and visualization of biological networks [12] BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks [13] Network thinking in ecology and evolution [14] DIP: The database of interacting proteins [15] IsoRankN: Spectral methods for global alignment of multiple protein networks [16] SMETANA: Accurate and scalable algorithm for probabilistic alignment of large-scale biological networks [17] Accurate multiple network alignment through context-sensitive random walk [18] BEAMS: Backbone extraction and merge strategy for the global many-to-many alignment of multiple PPI networks [19] Global multiple protein-protein interaction network alignment by combining pairwise network alignments

Copyright© 2019, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.

www.SID.ir

Archive of SID

چارچوب PRAF برای هم‌ترازی سراسری دو شبکه برهم‌کنش پروتئین- پروتئین

زهرا قربانعلی MSc
گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران
فاطمه زارع‌میرک‌آباد* PhD
گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده
به مجموعه ای از ماکرومولکول‌ها که در سلول با یکدیگر دارای تعامل هستند و عمل زیستی خاصی را انجام می‌دهند، شبکه زیستی گفته می‌شود. ناهنجاری تنها در یک مولکول اتفاق نمی‌افتد بلکه شبکه زیستی مربوط به آن را نیز درگیر می‌کند. برای شناسایی صحیح و جامع عوامل درگیر در یک بیماری باید از مقایسه بین شبکه‌های زیستی استفاده نمود. در این راستا، مسایل هم‌ترازی محلی و سراسری شبکه‌های برهم‌کنش پروتئین- پروتئین تعریف شد. با توجه به NP- کامل‌بودن مساله هم‌ترازی سراسری، الگوریتم‌های غیرقطعی مختلفی برای حل این مساله ارایه شده است. الگوریتم NetAl در این سال‌های اخیر به‌عنوان یک روش کارآمد برای حل این مساله شناخته شده است. گرچه این الگوریتم توانایی هم‌ترازی دو شبکه را با سرعت مناسبی دارد ولی ویژگی‌های زیستی را برای این منظور در نظر نمی‌گیرد. در این کار قصد داریم یک چارچوب جدید برای مساله هم‌ترازی سراسری شبکه‌های پروتئین- پروتئین به نام PRAF ارایه دهیم که با استفاده از این الگوریتم، نرم‌افزار BINGO و مفهوم هستی‌شناسی ژن، موجب بهبود نتایج الگوریتم NetAl شود.
کلیدواژه‌ها: بیوانفورماتیک، شبکه برهم‌کنش پروتئین- پروتئین، هم‌ترازی، هستی‌شناسی ژن، خوشه‌بندی

تاریخ دریافت: ۱۳۹۶/۱۱/۱۵
تاریخ پذیرش: ۱۳۹۶/۱۲/۱۲
*نویسنده مسئول: f.zare@aut.ac.ir

مقدمه

در زیست‌شناسی سیستمی، از مفهوم شبکه برای نمایش برهم‌کنش بین ماکرومولکول‌های زیستی مختلف استفاده می‌شود. ساختار و پویایی شبکه‌های زیست‌شناسی از تعامل بین مولکول‌ها در سلول منشا می‌گیرد. عملکردهای زیستی با مشارکت جمعی از ترکیبات سلولی مانند DNA، RNA، پروتئین‌ها و سایر مولکول‌های کوچک صورت می‌گیرد. شبکه‌های برهم‌کنش پروتئین- پروتئین که به اختصار PPI نامیده می‌شوند، به معنای ارتباط فیزیکی بین پروتئین‌هایی است که در اثر نیروهای الکترواستاتیکی یا رویدادهای شیمیایی به وجود می‌آید. معمولاً این برهم‌کنش‌ها به‌عنوان ارتباط فیزیکی یا تعامل بین پروتئین‌هایی که در یک سلول از یک موجود زنده در زمینه زیست‌مولکولی خاصی رخ می‌دهد، شناخته شده‌اند[1]. امروزه از روش‌های با بازدهی بالا مثل غربالگری دو هیبریدی[2] و تکنیک همکاری رسوب ایمونوتراپی به‌همراه طیف‌سنجی جرمی[3] برای شناسایی شبکه‌های PPI استفاده می‌شود. تجزیه و تحلیل مقایسه‌ای این شبکه‌ها می‌تواند در تشخیص مولفه‌های حفاظت‌شده بین شبکه‌ها و شناسایی نقش‌های عملکردی آنها کمک کند. با افزایش روزافزون شبکه‌های پروتئینی، اهمیت ارایه روش‌ها و ابزارهای کارآمد برای هم‌ترازی شبکه‌ها بیش‌تر می‌شود. از آنجایی که مساله هم‌ترازی شبکه‌ها، یک مساله NP-کامل است[4]، ارایه یک الگوریتم اکتشافی قابل اعتماد یکی از مهم‌ترین چالش‌ها در هم‌ترازی شبکه‌ها است.

مشابه با مساله هم‌ترازی توالی‌ها، روش‌های هم‌ترازی شبکه‌ها را می‌توان به دو نوع سراسری و محلی تقسیم نمود. در هم‌ترازی سراسری تمام رئوس شبکه‌ها به یکدیگر نگاشت شده ولی در هم‌ترازی محلی تنها زیرشبکه‌های مشابه با یکدیگر تراز می‌شوند.

یکی از الگوریتم‌های هم‌ترازی محلی که در این مقاله به آن اشاره شده، روش الگوریتم حریصانه و تصادفی برای همترازی محلی بهینه چندین شبکه برهم‌کنش (GASOLINE) [5] است. این الگوریتم شبکه‌های پروتئینی را براساس نمونه برداری تکراری و استراتژی حریصانه به‌صورت محلی هم‌تراز می‌کند. در این روش از روش دانه‌کاشتن و گسترش استفاده می‌شود تا کمپلکس‌های مشترک بین یک مجموعه از شبکه‌های PPI استخراج شود.

یکی از اولین الگوریتم‌ها در مساله هم‌ترازی سراسری شبکه‌ها، IsoRank [6] است. این الگوریتم از فلسفه‌ای شبیه به Google's PageRank استفاده می‌کند. به این صورت که اگر همسایگان دو راس به‌صورت مناسبی تطابق داشته باشند، آن دو راس بر هم نگاشت می‌شوند. براساس همین فرضیه، مساله هم‌ترازی تبدیل به مساله یافتن مقدار ویژه می‌شود. الگوریتم BinAligner [7] دو شبکه را براساس میزان شباهت توالی‌ها، یک‌همسایگی و چندهمسایگی با یکدیگر تراز می‌کند و در ادامه مساله به کمک الگوریتم‌های بهینه‌سازی حل می‌شود. اما یکی از معایب مهم این الگوریتم، عدم توانایی هم‌ترازی شبکه‌های با سایز بزرگ است. برای حل این چالش، الگوریتم NetAl [8] براساس یک مدل حریصانه دو شبکه را به‌صورت توپولوژیکی هم‌تراز می‌کند. هر چند این الگوریتم، از نظر سرعت، عملکرد خوبی دارد و به‌سادگی در دسترس است ولی یکی از نقاط ضعف آن درنظرنگرفتن ویژگی‌های زیستی است.

هدف اصلی ما در این مقاله افزودن ویژگی‌های زیستی به الگوریتم NetAl است. در این راستا یک چارچوب جدید به نام چارچوب هم‌ترازی شبکه‌های برهمکنش پروتئین- پروتئین (PRAF) معرفی می نماییم که با درنظرگرفتن ویژگی‌های زیستی دو شبکه، آنها را توسط الگوریتم NetAl هم‌تراز کند. یکی از مشخصات مهم زیستی، مفهوم هستی‌شناسی ژن‌ها است، به همین دلیل ما از الگوریتم BINGO استفاده می‌نماییم تا هر شبکه براساس این ویژگی خوشه‌بندی شود. خروجی این مرحله یک شبکه جدید از کلاسترها است که سایز کمتری از نظر راس و یال نسبت به دو شبکه ورودی دارد. در ادامه، برای هم‌ترازی این کلاسترها از نرم‌افزار GASOLINE استفاده می‌نماییم تا دو شبکه‌ی حاصل از مرحله قبل براساس مفهوم هستی‌شناسی ژن هم‌تراز شود. در پایان، الگوریتم NetAl روی هر جفت از کلاسترهایی که GASOLINE هم‌تراز نموده، اجرا می‌شود تا تطبیق رئوس داخلی آنها مشخص شود. چارچوب ارایه‌شده روی شبکه‌های مربوط به انسان، مخمر و موش اجرا و نتایج آن با الگوریتم NetAl مقایسه می‌شود. نتایج نشان می‌دهند که درنظرگرفتن ویژگی‌های زیستی موجب بهبود نتایج الگوریتم NetAl می‌شود.

مواد و روش‌ها

هر شبکه PPI با n پروتئین و m برهم‌کنش به‌صورت گراف $G = (V, E)$ تعریف می‌شود به‌طوری که هر $v_i \in V = \{v_1, v_2, \dots, v_n\}$، نماینده یکی از پروتئین‌های موجود در شبکه است و هر برهم‌کنش بین دو پروتئین v_i و v_j به‌صورت $(v_i, v_j) \in E = \{e_1, e_2, \dots, e_m\}$، نشان داده می‌شود.

همسایگی یک پروتئین از شبکه $G = (V, E)$، مجموعه پروتئین‌هایی است که با پروتئین مورد نظر برهم‌کنش دارند. به عبارت دیگر، همسایگی پروتئین $v_i \in V$ به‌صورت زیر تعریف می‌شود:

www.SID.ir

Archive of SID

$$N(v_i) = \{v_j | (v_i, v_j) \in E\}$$

تعریف مساله هم‌ترازی دو شبکه زیستی

همان‌طور که در مقدمه نیز اشاره شد دو نوع مساله برای هم‌ترازی شبکه‌های زیستی، هم‌ترازی سراسری و محلی، تعریف می‌شود. هدف ما در این مقاله ارایه یک چارچوب مناسب برای هم‌ترازی سراسری دو شبکه پروتئینی است.

در این مساله، ورودی دو شبکه PPI، $G_1 = (V, E)$ و $G_2 = (U, F)$ هستند به‌طوری که $|V| \geq |U|$ و خروجی تابعی مانند f است به‌طوری که $f: V \to U$، تابع f هر $v_i \in G_1$ را به $u_j \in G_2$ نگاشت می‌کند. این تابع یک‌به‌یک است و هر راس از شبکه اول تنها به یک راس از شبکه دوم نگاشت می‌شود.

شرح چارچوب PRAF

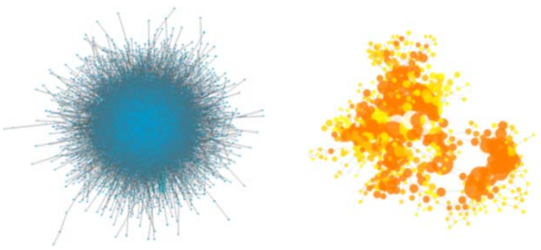
همان‌طور که اشاره شد، مساله هم‌ترازی سراسری، یک مساله مهم و چالش‌برانگیز در زیست‌شناسی محاسباتی است. اکثر الگوریتم‌های ارایه‌شده برای این مساله، از یک زیرگراف اولیه برای هم‌ترازی شبکه‌ها استفاده می‌کردند. واضح است که هر چه شبکه بزرگ‌تر باشد، روش انتخاب زیرگراف اولیه پیچیده‌تر می‌شود، چرا که اگر زیرگراف اولیه درست انتخاب نشود الگوریتم از مسیر اصلی هم‌ترازی منحرف می‌شود. به این منظور چارچوب PRAF معرفی می‌شود.

در ادامه گام‌های اصلی چارچوب PRAF که به شرح زیر هستند، به‌صورت دقیق‌تر بررسی می‌شوند:

۱) خوشه‌بندی شبکه‌ها
۲) هم‌ترازی محلی خوشه‌ها
۳) نگاشت داخلی خوشه‌های ترازشده
۴) هم‌ترازی پروتئین‌های باقی‌مانده
۵) هم‌ترازی سایر پروتئین‌ها به کمک توپولوژی گراف

خوشه‌بندی شبکه‌ها

برای خوشه‌بندی شبکه‌ها، راهکارها و الگوریتم‌های مختلفی وجود دارد که هر کدام از یک استراتژی استفاده می‌کنند. از جمله آنها می‌توان به Cluster Maker [9]، ClusterOne [10] و CytoCluster [11] اشاره کرد. هر سه روش شبکه را با توجه به توپولوژی گراف، خوشه‌بندی می‌کنند. ولی با توجه به این که شبکه‌های مورد بررسی، زیستی هستند، به نظر می‌رسد که به‌کاربردن اطلاعات هستی‌شناسی (GO) ژن برای خوشه‌بندی شبکه‌ها مناسب‌تر باشد. BINGO [12] ابزاری است که از آن برای دسته‌بندی یک شبکه زیستی براساس GO استفاده می‌شود. این نرم‌افزار، غالب عملکرد یک مجموعه ژن داده‌شده را روی درخت سلسله‌مراتب GO شناسایی می‌کند و خروجی این نگاشت به‌صورت یک گراف به نمایش در می‌آید. به عبارتی، این برنامه شبکه‌ها را با استفاده از ویژگی GO، براساس سه عملگر اصلی جز سلولی، عملکرد مولکولی و فرآیند زیستی به چندین خوشه تقسیم می‌نماید. در شکل ۱ نمایی از شبکه پروتئینی مخمر دیده می‌شود.



شکل ۱) تصویر سمت چپ شبکه خوشه‌بندی‌نشده و تصویر سمت راست خوشه‌بندی‌شده را نشان می‌دهد

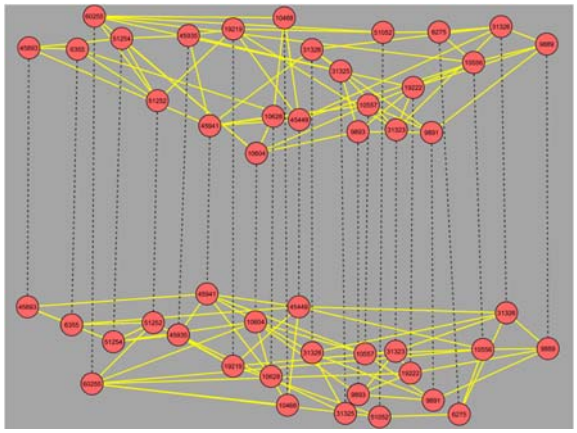
ورودی این برنامه یک فایل حاوی اطلاعات شبکه است به‌طوری که، هر سطر آن نشانگر یک جفت برهم‌کنش پروتئین- پروتئین از شبکه PPI است. خروجی این برنامه شامل دو فایل است. فایل اول حاوی اطلاعاتی نظیر نام هستی‌شناسی مربوطه، توصیف عملکرد این خوشه و مجموعه پروتئین‌هایی است که در این خوشه حضور دارند و فایل دوم ارتباط این خوشه‌ها را براساس درخت GO به‌صورت یک شبکه، که رئوس آنها خوشه‌ها هستند، نشان می‌دهد و GO-id مربوطه به‌عنوان نام آن راس قرار داده می‌شود.

به عبارت دیگر فایل دوم خروجی را می‌توان به‌صورت گراف $T_i = (V_T, G_T)$ برای هر کدام از شبکه‌ها نشان داد. به‌طوری که، $V_T = \{t_1, t_2, \ldots, t_n\}$ و (t_i, t_j) یک یال از گراف T_i را نشان می‌دهد، اگر BINGO [12] بین آنها براساس درخت GO ارتباطی تشخیص داده باشد. این گراف به‌عنوان ورودی به مرحله بعد وارد می‌شود.

هم‌ترازی محلی خوشه‌ها

همان‌طور که گفته شد، پیداکردن زیرگراف اولیه برای شروع هم‌ترازی یکی از مراحل مهم حل این مساله است. برای این امر پس از خوشه‌بندی شبکه‌ها به کمک BINGO [12]، خروجی این نرم‌افزار به‌صورت یک گراف، به نرم‌افزار GASOLINE [5] داده می‌شود. همان‌طور که در توضیح GASOLI0NE در مقدمه اشاره کردیم، این الگوریتم شبکه‌های PPI را به‌صورت محلی تراز می‌کند. در شکل ۲ نمایی از خروجی گرافیکی این برنامه نمایش داده شده است.

نرم‌افزار GASOLINE، اطلاعات شباهت خود را از دو روش به دست می‌آورد: ۱) داشتن فایل حاوی مقادیر امتیاز شباهت توالی بین هر جفت راس از شبکه‌هایی که تراز می‌شوند؛ ۲) تنها با درنظرگرفتن نام پروتئین‌ها. از آنجایی که نام پروتئین‌ها در این دو شبکه GO-id مربوط به خوشه است، معنادار است. در این نوع شباهت، نرم‌افزار خود را ملزم به نگاشت رئوس هم‌نام و دارای ارتباط کمپلکسی می‌نماید که در اینجا رئوس هم‌نام نشان‌دهنده یک نوع عملکرد خاص زیستی هستند. این مجموعه نگاشت یک دانه اولیه مناسب برای نرم‌افزار GASOLINE است تا آن را به یک هم‌ترازی مناسب گسترش دهد. سپس این الگوریتم به کمک توپولوژی شبکه یک هم‌ترازی محلی مناسب پیدا می‌کند.



شکل ۲) هم‌ترازی محلی شبکه‌های خوشه‌بندی‌شده

نگاشت داخلی خوشه‌های ترازشده

پس از نگاشت خوشه‌ها، آنها را براساس سایز، به‌صورت نزولی مرتب می‌نماییم. با توجه به این که ممکن است بعضی پروتئین‌ها دارای چندین عملکرد باشند، بنابراین امکان حضور آنها در چندین خوشه وجود دارد. برای این که بتوان هر پروتئین را فقط برای یک خوشه

www.SID.ir

Archive of SID

معرفی کرد، این پروتئین‌ها را از خوشه‌هایی که سایز کمتری دارند حذف می‌کنیم و آنها را تنها در یک خوشه در نظر می‌گیریم.
در ادامه، شبکه خوشه‌ها از روی شبکه اصلی بیرون کشیده می‌شود. سپس این شبکه‌ها توسط نرم‌افزار NetAl [8] به یکدیگر تراز می‌شوند. خروجی این مرحله، یک مجموعه از جفت پروتئین‌های ترازشده به‌صورت زیر است:

$$A = \{(v,u) | v \in V_A , u \in U_A\}$$

به‌طوری که V_A مجموعه رئوس ترازشده از شبکه G_1 و U_A مجموعه رئوس ترازشده از شبکه G_2 هستند.

هم‌ترازی پروتئین‌های باقی مانده

در مرحله قبل، پروتئین‌های درون خوشه‌ها تراز شدند. در این مرحله رئوس ترازشده، از گراف اصلی حذف می‌شوند. در این حالت، ابتدا تمام رئوس ترازشده از هر دو شبکه اصلی حذف و سپس دو شبکه جدید توسط NetAl [8] تراز می‌شوند. از آنجایی که درجه بعضی رئوس برابر یک است و ممکن است همسایه‌شان در ابتدا تراز شده باشد، این رئوس تنها، در ساختار شبکه جدید قرار نمی‌گیرند. بنابراین برای هم‌ترازی رئوس باقی‌مانده از شباهت توپولوژیکی استفاده می‌شود. در نظریه گراف و بررسی شبکه، مشخصه مرکزیت، مهم‌ترین رئوس یک گراف را تعیین می‌کند. از مهم‌ترین حالت‌های مرکزیت می‌توان به مرکزیت میانی و مرکزیت نزدیکی اشاره کرد[13]. مرکزیت نزدیکی به این صورت تعریف می‌شود که بین تمام رئوس گراف‌های همبند یک متریک فاصله طبیعی به‌وسیله طول کوتاه‌ترین مسیر آنها تعریف شده است. دوری راس x به‌صورت مجموع فاصله آن از تمام رئوس است و میزان نزدیکی عددی مقابل دوری تعریف می‌شود و به‌صورت زیر است:

$$C(x) = \frac{1}{\sum_y d(y,x)}$$

بنابراین رئوس مرکزتر، مجموع فاصله کمتری از تمام رئوس دارد. رئوسی که امتیاز توپولوژیکی نزدیک‌تری دارند به‌صورت حریصانه به یکدیگر تراز می‌شوند.

یافته‌ها

در این بخش، ابتدا پایگاه داده‌ای که مورد استفاده قرارگرفته، معرفی می‌شود، سپس تعدادی از معیارهای ارزیابی الگوریتم‌های هم‌ترازی دو شبکه شرح داده می‌شوند. درنهایت چارچوب PRAF با الگوریتم NetAl مقایسه می‌شود.

پایگاه داده

پایگاه داده استفاده‌شده در این مقاله، DIP [14] نام دارد. پایگاه داده DIP یک پایگاه داده زیستی است که برهم‌کنش بین پروتئین‌ها را به‌صورت تجربی و آزمایشگاهی شناسایی می‌کند. داده‌های ذخیره‌شده در DIP هم به‌صورت دستی، توسط متخصص و هم به‌صورت خودکار، با استفاده از روش‌های محاسباتی، گزینش می‌شوند. این پایگاه داده در سال ۲۰۰۲ منتشر شد و شامل هشت گونه زیر است:
دروزوفیلا ملانوگاستر (مگس میوه؛ *D. melanogaster*)، *ساکارومایسس سرویزیه* (مخمر نان؛ S. cerevisia)، *اشریشیا کلی* (*E. coli*)، انسان (هومو ساپینس)، سینورابتیدیس الگانس (*C. elegans*)، *موس موسکولوس* (*M. musculus*)، *هیلوباکتر پیلوری* (*H. pylori*)، *راتوس نروژیکوس* (*R. norvegicus*)
یکی از ویژگی مهم این پایگاه داده، دردسترس‌بودن ساختار اول (توالی) پروتئین‌های موجود در شبکه‌های DIP است. این توالی‌ها هنگام محاسبه کارآیی الگوریتم‌های هم‌ترازی مورد استفاده قرار می‌گیرند.

معیارهای ارزیابی

در ادامه، معیارهای صحت یال و میانگین نرمال‌شده امتیاز شباهت توالی برای ارزیابی هم‌ترازی دو شبکه معرفی می‌شود:

صحت یال

برای دو شبکه $G_1 = (V,E)$ و $G_2 = (U,F)$ امتیاز صحت یال به‌صورت زیر محاسبه می‌شود:

$$EC = \frac{|EdgeCor|}{\min\{|F|,|E|\}} \times 100\ \%$$

به‌طوری که:

$$EdgeCor = \{ (v,v') \in E | (f(v),f(v')) \in F \}$$

به‌طوری که f(v) نشان‌دهنده راس نگاشت‌شده معادل v در گراف G_2 است. به عبارتی EC، درصد تعداد یال‌هایی از شبکه اول است که به یالی در شبکه دوم تراز می‌شوند. مقدار بالای این امتیاز نشان می‌دهد که دو شبکه از نظر توپولوژیکی به یکدیگر شباهت دارند.

میانگین نرمال‌شده امتیاز شباهت توالی

برای محاسبه میزان شباهت توالی هم‌ترازی نهایی از این معیار استفاده می‌شود که به شکل زیر قابل محاسبه است:

$$ANSS = \sum_{v \in V} \frac{SS(v,f(v))}{((SS(v,v) \times SS(f(v),f(v)))^{-\frac{1}{2}}}$$

به‌طوری که f(v) نشان‌دهنده راس نگاشت‌شده معادل v در گراف G_2 است و SS(v, f(v)) میزان شباهت توالی دو پروتئین $v \in V$ و $f(v) \in U$ را نشان می‌دهد. هر چه این امتیاز بالاتر باشد بیان‌کننده آن است که رئوس نگاشت‌شده، از نظر زیستی مشابه هستند و هر چه این امتیاز کم‌تر باشد مشخص‌کننده آن است که رئوس همولوگ به‌خوبی به یکدیگر نگاشت نشده‌اند.

مقایسه چارچوب PRAF و الگوریتم NetAL

چارچوب PRAF روی یک دستگاه با پردازنده Intel® core™ i5-4200U CPU @ 1.60 GHz 2.29 GHz و یک حافظه دسترسی تصادفی (RAM) ۸گیگابایت اجرا می‌شود. از سه شبکه‌های انسان، مخمر و موش که اطلاعات آنها در جدول ۱ ذکر شده است، برای این کار استفاده کردیم.
در جدول‌های ۲ و ۳ میزان امتیاز صحت یال نشان داده شده است. الگوریتم NetAl تنها براساس میزان شباهت توپولوژیکی عمل می‌کند، به همین دلیل امتیاز بهتری در معیار صحت یال دارد.
نتایج معیار ANSS در جدول‌های ۴ و ۵ نشان داده شده است. همان‌طور که شرح داده شد، هر چه امتیاز ANSS بالاتر باشد، رئوس نگاشت‌شده از نظر زیستی شباهت بیشتری دارند. مقادیر ANSS روی شبکه‌ها مشخص می‌کند که این چارچوب از نظر زیستی عملکردی بهتر از NetAl دارد.
گرچه قبل از اجرای NetAl دو الگوریتم GASOLINEوBINGO اجرا می‌شوند ولی سرعت بالای این نرم‌افزارها و کمترکردن پیچیدگی شبکه‌هایی که به‌عنوان ورودی به NetAl می‌دهند، اختلاف چشمگیری بین زمان اجرای الگوریتم NetAl و چارچوب PRAF ایجاد نمی‌شود.

جدول ۱) اطلاعات شبکه‌های مورد استفاده

نام شبکه	نام علمی	تعداد رئوس	کمترین درجه رئوس	بیشترین درجه رئوس	تعداد یال‌ها
مخمر	*ساکارومایسس سرویزیه*	۴۶۴۵	۱	۲۴۱	۲۱۷۹۵
انسان	هوموساپینس	۳۹۹۱	۱	۲۰۴	۶۱۱۰
موش	*موس موسکولوس*	۱۲۴۲	۱	۳۵	۱۳۷۸

www.SID.ir

Archive of SID

جدول ۲) امتیاز صحت یال برای هم‌ترازی دو شبکه براساس PRAF

نام شبکه‌ها	EC
انسان- مخمر	%۳۹/۲۱۴
موش- مخمر	%۴۸/۳۳۱
موش- انسان	%۵۶/۷۴۸

جدول ۳) امتیاز صحت یال برای هم‌ترازی دو شبکه براساس الگوریتم NetAl

نام شبکه‌ها	EC
انسان- مخمر	%۵۳/۷۶۴
موش- مخمر	%۵۰/۴۳۵
موش- انسان	%۶۴/۶۵۸

جدول ٤) امتیاز ANSS برای هم‌ترازی دو شبکه براساس PRAF

نام شبکه‌ها	ANSS
انسان- مخمر	۱۱/۳۷۵
موش- مخمر	۱۹/۲۳۰
موش- انسان	۸/۹۷۷

جدول ۵) امتیاز ANSS برای هم‌ترازی دو شبکه براساس NetAl

نام شبکه‌ها	ANSS
انسان-مخمر	۱۰/۹۸۳
موش-مخمر	۱٤/۳۸۲
موش-انسان	۳/۲۰٤

بحث

هدف اصلی در این مقاله ارایه چارچوبی برای بهینه‌نمودن و افزودن جنبه زیستی به الگوریتم NetAl است. در مرحله چهارم این چارچوب، رئوس کلاسترهای هم‌ترازشده از ساختار دو شبکه حذف می‌شود و درنهایت رئوس باقی‌مانده در دو شبکه جدید با الگوریتم NetAl هم‌تراز می‌شود. با توجه به این که این فرآیند موجب تغییراتی در توپولوژی شبکه می‌شود، هدف ما در آینده ایجاد تغییراتی در ورودی الگوریتم NetAl است به‌طوری که رئوس مشخص‌شده در ورودی از هر شبکه را در هم‌ترازی دخالت دهد. به‌نظر می‌رسد که این تغییر در الگوریتم، کارآیی PRAF را در هم‌ترازی دو شبکه افزایش دهد زیر دیگر نیازی به حذف رئوس شبکه در مرحله چهارم PRAF نداریم.

با توجه به پیشرفت روش‌های بیوتکنولوژی، شبکه‌های برهم‌کنش پروتئین- پروتئین بسیاری از گونه‌ها در دسترس قرار گرفته است. دردسترس‌بودن شبکه‌های برهم‌کنش پروتئینی گونه‌های مختلف، موجب توسعه روش‌های محاسباتی برای شناخت و تفسیر عملکرد پروتئین‌ها شده است. یکی دیگر از این روش‌ها که امکان بررسی همزمان چندین شبکه را در اختیار ما قرار می‌دهد، هم‌ترازی سراسری چندگانه شبکه‌های پروتئینی است. الگوریتم IsoRankN [15] اولین الگوریتم هم‌ترازی سراسری چند شبکه برهم‌کنش پروتئین- پروتئین است که از ایده Google's PageRank بهره می‌برد. این الگوریتم از روش تقسیم محلی گراف برای تشکیل خوشه‌های هم‌ترازی استفاده می‌کند. الگوریتم‌های SMETANA [16] و CSRW [17] از گام تصادفی برای محاسبه‌ی احتمال تطبیق جفت گره‌های شبکه‌های متفاوت استفاده می‌نمایند. الگوریتم BEAMS [18] تنها شباهت توالی پروتئین‌ها برای ساخت خوشه‌ها را در نظر می‌گیرد. الگوریتم SMAL [19] نتیجه الگوریتم‌های هم‌ترازی سراسری دو شبکه را به هم‌ترازی سراسری چند شبکه گسترش می‌دهد. تمام این الگوریتم‌ها از یک تابع هزینه برای محاسبه شباهت توالی و ساختاری پروتئین‌ها در گونه‌های مختلف استفاده می‌کنند. همچنین در بسیاری از این الگوریتم‌ها، شباهت گره‌ها در شبکه‌های مختلف با استفاده از روش‌های احتمالاتی محاسبه می‌شود. این امر موجب افزایش پیچیدگی محاسباتی الگوریتم‌ها می‌شود.

در آینده ایده مطرح‌شده در این مقاله را می‌توان برای هم‌ترازی چندین شبکه زیستی گسترش داد. به این ترتیب که ابتدا تمام شبکه‌ها را به کمک نرم‌افزار BINGO خوشه‌بندی نموده و سپس خوشه‌ها را به‌وسیله GASOLINE هم‌تراز کرد. در ادامه پروتئین‌های تشکیل‌دهنده خوشه‌ها را دوبه‌دو تراز نمود و ماتریسی از امتیاز هم‌ترازی آنها ساخت. سپس مشابه‌ترین جفت شبکه را با یکدیگر ترکیب کرد و یک شبکه جدید تشکیل داد و درنهایت شبکه‌ها را با شبکه جدید هم‌تراز نمود. این فرآیند تا نگاشت تمام پروتئین‌ها ادامه می‌یابد. به این ترتیب می‌توان از الگوریتم NetAl که دارای سرعت بالایی است، استفاده نمود تا چندین شبکه را هم‌تراز کرد.

نتیجه‌گیری

در این مقاله برای حل مساله هم‌ترازی سراسری دو شبکه برهم‌کنش پروتئین- پروتئین چارچوبی به نام PRAF ارایه نمودیم. چارچوب PRAF از سه برنامه اصلی NetAl، GASOLINE، BINGO و مفهوم هستی‌شناسی ژن برای هم‌ترازی دو شبکه بهره می‌برد. هدف اصلی PRAF افزودن مفهوم زیستی به الگوریتم همترازی NetAl است.

تشکر و قدردانی: نویسندگان مقاله مراتب تشکر و قدردانی خود را از دانشگاه امیرکبیر ابراز می‌دارند.

تاییدیه‌اخلاقی: موردی از سوی نویسندگان گزارش نشده است.

تعارض منافع: موردی از سوی نویسندگان گزارش نشده است.

سهم نویسندگان: زهرا قربانعلی (نویسنده اول)، نگارنده مقدمه/روش‌شناس/پژوهشگر اصلی (۵۰%)؛ فاطمه زارع‌میرک‌آباد (نویسنده دوم)، روش‌شناس/پژوهشگر اصلی/تحلیلگر آماری/نگارنده بحث (۵۰%)

منابع مالی: موردی از سوی نویسندگان گزارش نشده است.

منابع

1- Rivas DL, Fontanillo C. Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. PLoS Computat Biol. 2010;6(6):e1000807.

2- Young KH. Yeast two-hybrid: so many interactions, (in) so little time. Biol Reprod. 1998;58(2):302-11.

3- Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

4- Lathrop RH. The protein threading problem with sequence amino acid interaction preferences is NP-complete. Protein Eng. 1994;7(9):1059-68.

5- Micale G, Pulvirenti A, Giugno R, Ferro A. GASOLINE: a Greedy and Stochastic algorithm for Optimal Local multiple alignment of Interaction Networks. Plos One. 2014;9(6):e98750.

6- Singh R, Xu J., Berger B. Global alignment of multiple protein interaction networks with application to functional orthology detection. PNAS. 2008;105(35);12763-12768.

7- Yang J, Li J, Grunewald S, Wan XF. BinAligner: A heuristic method to align biological networks. BMC Bioinformatics. 2013;14(Suppl 14):S8.

www.SID.ir

Archive of SID

8- Neyshabur B, Khadem A, Hashemifar S, Arab S. NETAL: A new graph-based method for global alignment of protein-protein interaction networks. Bioinformatics. 2013;29(13):1654-62.
9- Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G et al. Clustermaker: A multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics. 2011;12:436.
10- Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods. 2012;9:471-2.
11- Li M, Tang Y, Li D, Wu F, Wang J. CytoCluster: A cytoscape plugin for cluster analysis and visualization of biological networks. Int J Mol Sci. 2017;18(9):E1880.
12- Maere S, Heymans K, Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448-9.
13- Proulx SR, Promislow DE, Phillips PC. Network thinking in ecology and evolution. Trends Ecol Evol. 2005;20(6):345-53.
14- Xenarios I, Rice WD, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: The database of interacting proteins. Nucleic Acids Res. 2000;28(1):289-91.
15- Liao CS, Lu K, Baym M, Singh R, Berger B. IsoRankN: Spectral methods for global alignment of multiple protein networks. Bioinformatics. 2009;25(12:i253-i258.
16- Sahraeian M, Yoon BJ. SMETANA: Accurate and scalable algorithm for probabilistic alignment of large-scale biological networks. PloS One. 2015;8(7):e67995.
17- Jeong H, Yoon B. Accurate multiple network alignment through context-sensitive random walk. BMC Sys Biol. 2015;9(Suppl 1):S7.
18- Alkan F, Erten C. BEAMS: Backbone extraction and merge strategy for the global many-to-many alignment of multiple PPI networks. Bioinformatics. 2014;30(4):531-9.
19- Dohrmann J, Puchin J, Singh R. Global multiple protein-protein interaction network alignment by combining pairwise network alignments. BMC Bioinformatics. 2015;16(Suppl 13):S11.

www.SID.ir