مدل‌سازی شبکه‌های تنظیم ژنی: مدل‌های کلاسیک، اختلال بهینه برای شناسایی شبکه

رضا بیات[1*]، مهدی صادقی[2] و محمدرضا عارف[3]

[1] دانشکدهٔ مهندسی برق، دانشگاه یزد، یزد، ایران؛ و همکار ISSL §

[2] پژوهشکدهٔ زیست‌فناوری پزشکی، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری، تهران، ایران

[3] گروه مخابرات، دانشکدهٔ مهندسی برق، دانشگاه صنعتی شریف، تهران، ایران؛ و استاد وابستهٔ دانشگاه یزد

چکیده

ارتقای عمق و گسترهٔ درک ما از دانش زیست‌شناسی ملکولی، از یک سو امکان بهره‌برداری از آن را در توسعهٔ فناوری‌هایی مانند رمزگشایی فراهم ساخته است و از سوی دیگر، مداخله در سیستم ژنتیکی را امکان‌پذیر می‌سازد که نویدبخش آینده‌ای روشن برای علوم زیستی و پزشکی است. دست‌یابی به این هدف با مداخله در شبکهٔ تنظیم ژنی (GRN) امکان‌پذیر می‌شود؛ زیرا GRN کنترل‌کنندهٔ فعالیت‌های زیستی در سطح ملکولی است. در این مسیر، شناسایی GRN، شامل شناسایی مرز، ساختار و گره‌های شبکه اهمیت به‌سزایی دارد. در این مقاله به دو جنبهٔ ساختار و گره در مدل‌سازی و شناسایی GRN در شبکه‌های بزرگ (با بیش از پنجاه گره) پرداخته می‌شود. نخست محدودیت‌های کاربست مدل‌های احتمالاتی برای گره (ژن) مورد بررسی قرار می‌گیرد. همچنین محدودیت‌های کاربست مدل چند-درختی برای ساختار GRN مورد بررسی قرار می‌گیرد. در بخش اصلی مقاله، مسألهٔ شناسایی GRN با مدل بولی مورد بحث قرار گرفته و نشان داده می‌شود که برخلاف تصور معمول، آزمایش بهینه از دید شناسایی ساختار GRN، آزمایش تک‌اختلال است.

واژگان کلیدی: شبکهٔ تنظیم ژنی، مدل احتمالاتیِ ژن، مدل چنددرختیِ ساختار، مدل بولیِ ژن، آزمایش بهینهٔ اختلال.

Modeling gene regulatory networks: Classical models, optimal perturbation for identification of network

Reza Bayat[1*], Mehdi Sadeghi[2] & Mohammad Reza Aref[3]

[1] Department of Electrical Engineering, Yazd University, Yazd, Iran

[2] Medical Genetics Department, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

[3] Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran; and Yazd University, Yazd, Iran

Abstract

Deep understanding of molecular biology has allowed emergence of new technologies like DNA decryption. On the other hand, advancements of molecular biology have made manipulation of genetic systems simpler than ever; this promises extraordinary progress in biological, medical and biotechnological applications. This is not an unrealistic goal since genes which are regulated by gene regulatory networks (GRNs) are the core governors of life processes at the molecular level. In fact, manipulation of GRNs would be the ultimate strategy for optimal purposeful control of cell's life. GRNs are in charge of regulating the amounts of all the inter-cellular as well as intra-cellular molecular species produced all the time in all living organisms. Manipulation of a GRN requires comprehensive knowledge about nodes and interconnections. This paper deals with both aspects in networks having more than fifty nodes. In the first part of the paper, restrictions of probabilistic models in modeling node behavior are discussed, i.e.: 1) unfeasibility of reliably predicting the next state of GRN based on its current state, 2) impossibility of modelling logical relations among genes,

* Corresponding author

*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۳۹۷/۰۸/۰۱ • تاریخ پذیرش: ۱۳۹۷/۱۱/۰۶ • تاریخ انتشار: ۱۳۹۹/۰۶/۲۴ • نوع مطالعه: پژوهشی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]

and 3) scarcity of biological data needed for model identification. These findings which are supported by arguments from probability theory suggest that probabilistic models should not be used for analysis and prediction of node behavior in GRNs. Next part of the paper focuses on models of GRN structure. It is shown that the use of multi-tree models for structure for GRN poses severe limitations on network behavior, i.e. 1) increase in signal entropy while passing through the network, 2) decrease in signal bandwidth while passing through the network, and 3) lack of feedback as a key element for oscillatory and/or autonomous behavior (a requirement for any biological network). To demonstrate that, these restrictions are consequences of model selection, we use information theoretic arguments. At the last and the most important part of the paper we look into the gene perturbation experiments from a network-theoretic perspective to show that multi-perturbation experiments are not as informative as assumed so far. A generally accepted belief among researches states that multi-perturbation experiments are more informative than single-perturbation ones, i.e., multiple simultaneously applied perturbations provide more information than a single perturbation. It is shown that single-perturbation experiments are optimal for identification of network structure, provided the ultimate goal is to discover correct subnet structures.

Keywords: gene regulatory network (GRN), probabilistic model of gene, multi-tree model of GRN structure, Boolean model of gene, optimal perturbation experiment.

۱- مقدمه

درجهٔ بالای بهینگی در ساختار و عملکرد موجودات یا سیستم‌های زنده - از ساده‌ترین تا پیچیده‌ترین - این فکر را القا می‌کند که شناخت آنها می‌تواند راه‌گشای حل مسائل پیش روی بشر باشد. برای مثال توانایی اصلاح خطا در انتقال صفت‌ها از والدین به فرزندان توسط سیستم ژنتیکی بسیار بالا است. برای دست‌یابی به نرخ پایین خطا در کدهای مخابراتی، طول کد باید زیاد باشد، اما در سیستم‌های زیستی، انتقال اطلاعات صفت‌ها با استفاده از کدهایی انجام می‌شود که با وجود طول بسیار کمتر، نرخ خطای بسیار کمتری دارند [1].

هم‌افزایی نظریهٔ سیستم و زیست‌شناسی ملکولی، موجب پیدایش طیف وسیعی از پژوهش‌های میان‌رشته‌ای در دو شاخهٔ اصلی زیر شده است:

۱- پژوهش در زیست‌شناسی و شناخت اجزای سیستم ژنتیکی با هدف مداخلهٔ هدفمند در آن: شکل‌های ساده‌ای از مداخله در گذشته نیز برای بهسازی نژاد دام‌ها به‌کار رفته است.

۲- شناسایی سیستم‌های زیستی با هدف ابداع روش‌های نوین مهندسی: برای نمونه، رمزگشایی DNA [2] روشی کارامد است که در آن از ویژگی پردازش موازی در ملکول‌ها استفاده می‌شود تا فضای بسیار بزرگ جست‌وجو به‌سرعت پیموده شود. مقایسه‌ای بین روش‌های کدنگاری مخابراتی و ژنتیکی در [1] و [5] ارائه شده است. کاربردهایی جدیدتر در [3, 4] ارائه شده است.

ویژگی‌های زیستی سلول جاندار در واحدهایی به نام ژن حمل می‌شود. ملکول DNA حاوی تمامی ژن‌ها است و نسخه‌ای از آن در هر یک از سلول‌ها نگهداری می‌شود. DNA سیستم ذخیره‌سازی اطلاعات ژنتیکی است و انتقال اطلاعات صفت‌ها به نسل بعدی و همچنین بهره‌برداری از این اطلاعات را هدایت می‌کند [6].

شبکه‌های تنظیم بیان ژن[1] (GRN-ها) در سطوح مختلفی در پیکر جاندار در کار هستند و تنظیم فرایندهای زیستی را در سطح ملکولی به‌گونه‌ای انجام می‌دهند تا نوعی هماهنگی سراسری بین عملکرد ژن‌ها ایجاد شود [7]. مداخله در عملکرد GRN-ها به‌منظور کنترل فعالیت ژن‌ها، نویدبخش آینده‌ای روشن برای نظام سلامت با اهداف زیر است: الف) درمان بیماری‌های دارای منشأ ژنتیکی مانند اوتیسم در دوران جنینی [8]؛ ب) شخصی‌سازی پزشکی و طراحی دارو با درنظرگرفتن تفاوت‌های ژنتیکی [9]؛ پ) دست‌یابی به عمر طولانی با تشخیص و درمان بیماری‌های ژنتیکی و همچنین از طریق جایگزینی عضو آسیب‌دیده با نمونهٔ ساخته‌شده بر مبنای محتوای ژنتیکی فرد [10].

برای موفقیت در مداخله، به اطلاعات زیر نیاز است:

۱) محدوده یا مرز GRN

۲) ساختار ارتباطی بین گره‌ها (ژن‌ها) در GRN

۳) ویژگی‌های پردازشی گره‌های GRN

با دراختیارداشتن این اطلاعات و با اعمال مداخلهٔ مناسب، می‌توان رفتار مورد نظر را در شبکه القا کرد [7, 11].

موضوع مقالهٔ حاضر شناسایی GRN است؛ یعنی:

۱- شناسایی ساختار ارتباط بین گره‌ها

۲- شناسایی ویژگی‌های پردازشی گره‌ها

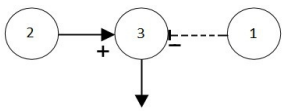
در پژوهش‌های پیشین به چند موضوع مرتبط با شناسایی GRN کمتر پرداخته شده است. برای نمونه، با آنکه وضعیت هر ژن در سطح ملکولی، تابعی منطقی از وضعیت ژن‌های مرتبط با آن است [7, 8]، مدل‌سازی و شناسایی

[1] gene expression regulation networks

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]

رابطهٔ منطقی بین ژن‌ها کمتر مورد توجه قرار گرفته و در بیشتر پژوهش‌ها، نقشهٔ تأثیرگذاری[1] ژن‌ها مانند آنچه در شکل (۱) نشان داده شده، استنتاج می‌شود [12-17] که نوعی گراف جهت‌دار برای نمایش رابطهٔ علت و معلولی بین ژن‌ها است. در نقشهٔ تأثیرگذاری شکل (۱)، دو ژن g_1 و g_2، بر ژن g_3 تأثیر می‌گذارند. در شبکه‌های تأثیرگذاری، اطلاعات قابل نمایش در این حد است: فعالیت ژن منبع یعنی g_1 تأثیری (در اینجا) کاهنده بر فعالیت ژن مقصد یعنی g_3 دارد؛ اما سازوکار منطقی این تأثیرگذاری مشخص نمی‌شود. در شبکه‌های تأثیرگذاری، به‌طورمعمول از مدل‌های احتمالاتی برای کمّی‌سازی تأثیر گره‌ها استفاده می‌شود.



(شکل-۱): مثالی از نقشهٔ تأثیرگذاری
(Figure-1): Sample Influence Map

محدودیتی مهم در مدل تأثیرگذاری آن است که پیش‌بینی فعالیت ژن در آن امکان‌پذیر نیست. در مثال بالا رابطهٔ خروجی با ورودی‌ها می‌تواند به یکی از دو شکل زیر باشد، اما مدل تأثیرگذاری برای آنها یکسان است:

$$g_3 = \overline{g_1} g_2 \qquad \text{(AND)} \qquad (۱)$$

$$g_3 = \overline{g_1} + g_2 \qquad \text{(OR)} \qquad (۲)$$

این دو رابطه فقط در پنجاه درصد از حالت‌های ورودی‌ها، خروجی یکسانی دارند و در پنجاه درصد دیگر، خروجی‌های آنها معکوس یکدیگر است. در قسمت ۲-۲ محدودیت‌های مدل تأثیرگذاری بر پایهٔ توصیف احتمالاتی بررسی می‌شود.

از سوی دیگر، به‌طورتقریبی در تمامی تحلیل‌های GRN، وجود بازخورد زیستی نادیده گرفته می‌شود [12-19]. محدودیت‌های اساسی ناشی از این فرض در مورد ساختار شبکه، موضوع قسمت ۲-۳ است. در بسیاری از پژوهش‌ها از ساختار بدون بازخورد همراه با مدل احتمالاتی استفاده می‌شود که محدودیت‌های هر دو مدل را در بر دارد [12-17].

مسألهٔ دیگری که در بخش پایانی مقاله مورد بررسی قرار می‌گیرد، مدل بولی GRN و کاربست آزمایش‌های اختلال ژنی[2] برای شناسایی ساختار مدل است. کاربست مدل بولی به‌عنوان مدلی کیفی برای مدل‌سازی GRN سابقه‌ای طولانی دارد [20, 21]، اما به‌تازگی در کاربردهای واقعی نیز مورد توجه قرار گرفته است [22]. در مدل بولی، گره یا ژن نوعی پردازش‌گر دودویی در نظر گرفته می‌شود که ورودی‌هایی را از تعدادی گره دیگر دریافت می‌کند. برای شناسایی شبکه از آزمایش‌های اختلال ژنی استفاده می‌شود [19]: با اعمالِ اختلال مصنوعی در عملکرد طبیعی ژن‌ها، تأثیرهای ناشی از آن مطالعه می‌شود. در پژوهش‌های پیشین، برخی مسائل مرتبط با مدل‌سازی بولی مورد توجه قرار نگرفته یا به‌درستی شناسایی و تبیین نشده است. در بخش سوم مقاله، با بررسی آزمایش‌های اختلال نشان داده می‌شود که نمی‌توان ساختار شبکه را فقط با استفاده از آزمایش‌های چنداختلال شناسایی کرد. نشان داده می‌شود که آزمایش تک‌اختلال بیشترین کارایی را دارد. بر اساس یافته‌های این پژوهش، باید نخست آزمایش‌های تک‌اختلال و سپس تعدادی آزمایش تکمیلی دو یا چنداختلال برای شناسایی شبکه انجام شود.

مقالهٔ حاضر چنین سازمان یافته است: در بخش دوم، نخست مبانی GRN و مدل‌سازی آن مرور می‌شود. سپس در دو قسمت ۲-۲ و ۲-۳ محدودیت‌های دو مدل پرکاربرد یعنی مدل احتمالاتی برای گره و مدل بدون بازخورد برای ساختار شبکه بررسی می‌شود. در بخش سوم، نخست مدل منطقی GRN و آزمایش اختلال معرفی و نشان داده می‌شود که آزمایش تک‌اختلال بیشترین کارایی را در شناسایی دارد.

۲- مدل‌های کلاسیک برای GRN

در این بخش، نخست مفاهیم پایه و همچنین روش‌های کلاسیک مدل‌سازی GRN معرفی و مرور می‌شود؛ سپس محدودیت‌های دو مدل کلاسیک برای گره و ساختار مورد بررسی قرار می‌گیرد.

۲-۱- شبکه‌های تنظیم ژنی

سلول کوچکترین واحد زنده است. انواع مختلف سلول‌ها در سامانه‌های مختلف زیستی در پیکر جاندار (حدود ۲۱۰ نوع در انسان [23])، فعالیت‌های زیستی متنوعی را انجام می‌دهند. تنظیم فعالیت‌ها در درون سلول و همچنین هماهنگ‌سازی بین سلول‌های مختلف، از طریق تبادل اطلاعات شیمیایی یا الکتریکی (در سلول‌های عصبی) انجام می‌شود [6]. تبادل پیوستهٔ اطلاعات در GRN، حیات سلولی

[1] influence map
[2] gene perturbation

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]
[DOI: 10.29252/jsdp.17.2.112]

را در شرایط مختلف تنظیم می‌کند [24]. در سلول‌های جاندار نسخهٔ یکسانی از ملکول DNA وجود دارد که در انسان حدود ۲۲ هزار ژن را حمل می‌کند و حاوی دستور ساخت حدود بیست هزار پروتئین و دو هزار عامل رونویسی (TF)[1] است [25]. پروتئین، عنوانی عمومی برای ملکول‌هایی است که بخش اعظم پوسته و محتوای سلول را در تمامی جانداران تشکیل می‌دهند.

اتصال TF به جایگاه اتصال در بخش راه‌اندازِ رونویسی[2] موجب افزایش یا کاهش فعالیت ژن می‌شود [26, 27]. TF نقش پیام‌رسانی و پروتئین (مانند هموگلوبین) نقشی زیستی (غیر پیام‌رسانی) مانند حمل اکسیژن را در GRN بر عهده دارد. به‌دلیل نیاز به تعداد بالای ملکول‌های پروتئین در سلول، خواندن اطلاعات ژن‌ها از روی "تنها" DNA-ی موجود در سلول باید به‌طور پیوسته و به تعداد دفعات زیادی انجام شود. بنابراین تنظیم فعالیت یا بیان ژن‌ها نیز فرایندی دائمی است.

انجام محاسبه در ژن نتیجه‌ای است از تأثیر شیمیایی مواد دریافت‌شده از سایر گره‌ها (ژن‌ها یا TF-ها) بر جایگاه اتصال ژن مورد نظر. برای مثال هنگامی که نشستن یک TF بر جایگاه اتصال یک ژن آن را روشن می‌کند، می‌توان فرض کرد که معادلهٔ ریاضی توصیف‌کنندهٔ عملکرد آن ژن تغییر می‌یابد. مهم‌ترین مدل‌های پیشنهادشده برای محاسبه در گره‌های زیستی عبارتند از [11, 28, 29]:

۱) مدل‌های دقیق
۲) مدل‌های احتمالاتی
۳) مدل‌های منطقی
۴) مدل‌های ترکیبی و شبکه‌های عصبی

در مدل‌های دقیق تحولات زمانی واکنش‌های شیمیایی مرتبط با فعالیت ژن با استفاده از قوانین فیزیکی-شیمیایی و زیستی مدل‌سازی و دستگاه معادلات دیفرانسیل حاصل تحلیل می‌شود [7, 30]. به‌علت تعداد زیاد پارامترها، این مدل‌ها حتی برای ده ژن پیچیدگی بالایی دارند [7]؛ زیرا: ۱) به تعداد زیادی اندازه‌گیری برای تخمین پارامترها نیاز است و ۲) توان پردازشی مورد نیاز برای حل معادلات و تخمین کمیت‌ها بالا است. به همین دلیل در سایر مدل‌ها از جمله در مدل احتمالاتی و مدل منطقی، فعالیت ژن‌ها به‌شکلی انتزاعی و برای مثال به شکل دوسطحه یا دودویی توصیف می‌شود تا مسألهٔ مورد نظر از دیدگاه محاسباتی ساده‌تر باشد.

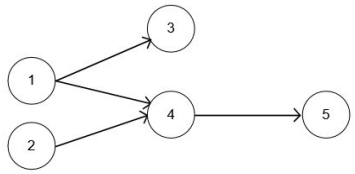
[1] transcription factor
[2] binding site in transcription initiation site

۲-۲- مدل احتمالاتی و محدودیت‌های آن

در حالت کلی مدل احتمالاتی برای کاربردهایی مناسب است که به ابزاری برای پردازش عدم قطعیت از نوع تصادفی‌بودن نیاز است، مانند مطالعهٔ اثر نوفه بر کانال‌ها و شبکه‌های مخابراتی [31]. در این قسمت نشان داده می‌شود که مدل احتمالاتی برای مدل‌سازی پردازش در ژن‌ها مناسب نیست.

از آغاز ظهور زیست‌شناسی محاسباتی، مدل احتمالاتی با استفاده از گراف جهت‌دار مارکف برای مطالعهٔ شبکه‌های زیستی پیشنهاد و به‌تدریج به مدلی پرکاربرد تبدیل شده است. در این مدل، بیان ژن پدیده‌ای احتمالی و بردار خروجی ژن‌ها برداری تصادفی فرض شده و توزیع احتمال بردار خروجی ژن، با استفاده از قاعدهٔ بیز[3] بر حسب توزیع شرطی بردارهای والد یا ورودی آن نوشته می‌شود. توزیع‌های احتمال باید از روی مشاهدات یا اندازه‌گیری‌های مفصل بیان هم‌زمان ژن‌ها به‌دست آید [11, 12, 14-17, 32].

شکل (۲) مثالی را از شبکهٔ مارکف نشان می‌دهد که در آن به‌علت مشاهدهٔ مکرر بیان ژن g_4 (هم‌زمان با بیان دو ژن g_1 و g_2)، فعالیت ژن g_4 به شکل مشروط به فعالیت دو ژن g_1 و g_2 مدل‌سازی شده است.



$$\boldsymbol{g} = (g_1, g_2, g_3, g_4, g_5)$$
$$p(\boldsymbol{g}) = p(g_5|g_4)p(g_4|g_1, g_2)p(g_3|g_1)p(g_1, g_2)$$

(شکل-۲): مثالی از شبکه‌های مارکف

(Figure-2): Example of a Markov Network Model

بیان ژن پدیده‌ای است که مدل‌سازی احتمالاتی برای آن مناسب نیست. برای مثال حتی اگر احتمال بیان ژن برابر ۰/۹ به‌دست آید، در مدل احتمالاتی نمی‌توان از روشن یا خاموش‌بودن آن در زمانی خاص اطمینان داشت یا بیان آن را در شرایطی مشخص، تضمین کرد.

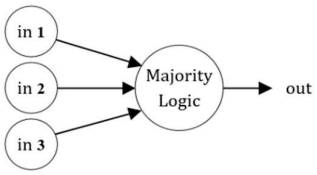
مشکل دوم مدل احتمالاتی، نیاز به حجم بالای داده برای استنتاج گراف شبکه است. بر خلاف کاربردهای مخابراتی، دسترسی به دادهٔ زیستی پرهزینه است. در مورد آزمایش‌های اختلال، دسترسی به دادهٔ فراوان به‌طوراساسی امکان‌پذیر نیست؛ زیرا تعداد کل آزمایش‌های اختلال محدود است.

[3] Bayes' rule

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]
[DOI: 10.29252/jsdp.17.2.112]

اما مشکل مهم مدل احتمالاتی، ناتوانی آن در توصیف رابطهٔ منطقی بین ژن‌ها است. این موضوع با استفاده از زیرشبکهٔ ژنتیکی ساده‌ای نشان داده می‌شود که منطق اکثریت[1] نام دارد. زیرشبکهٔ چهار-گرهی شکل (۳) را در نظر بگیرید. مسألهٔ منطق اکثریت چنین تعریف می‌شود:

out = 1; if majority of inputs are 1
out = 0; otherwise



(شکل-۳): زیرشبکهٔ منطق اکثریت
(Figure-3): Majority Logic Subnet

جدول صحت مسأله در جدول (۱) آورده شده است. تابع منطقی توصیف‌کنندهٔ خروجی عبارت است از:

$$g_{out} = \overline{g_1}g_2g_3 + g_1\overline{g_2}g_3 + g_1g_2g_3 + g_1g_2\overline{g_3} \qquad (۳)$$

معادلهٔ (۳) رابطهٔ منطقی بین گره‌ها را به‌طور کامل توصیف می‌کند.

(جدول-۱): جدول صحت زیرشبکهٔ منطق اکثریت
(Table-1): Truth Table for Majority Logic Subnet

index	g_3	g_2	g_1	g_{out}
0	0	0	0	0
1	0	0	1	0
2	0	1	0	0
3	0	1	1	1
4	1	0	0	0
5	1	0	1	1
6	1	1	0	1
7	1	1	1	1

در مدل احتمالاتی، رابطهٔ توزیع احتمال سه ورودی با توزیع احتمال خروجی با فرض استقلال ورودی‌ها چنین است:

$$p(out) = p(out|g_1,g_2,g_3)p(g_1,g_2,g_3) = p(out|g_1,g_2,g_3)p(g_1)p(g_2)p(g_3) \qquad (۴)$$

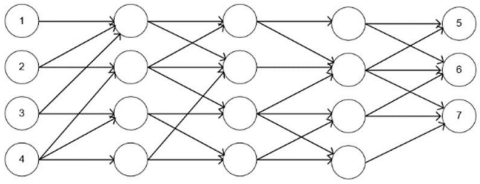
با استفاده از این توصیف نمی‌توان سازوکار یا منطق عملکرد جمعی یا هم‌افزایی ورودی‌ها برای ایجاد خروجی را توصیف کرد. درواقع از توصیف احتمالاتی نمی‌توان نتیجه گرفت که برای روشن‌شدن خروجی، چه تعدادی از ورودی‌ها باید روشن باشند.

رابطهٔ منطقی بین ژن‌ها به‌طورمعمول پیچیده‌تر از شمارش تعداد ورودی‌های فعال (روشن) است و مدل‌سازیِ سازوکارهای تعاملِ گروهی با استفاده از مدل احتمالاتی امکان‌پذیر نیست.

[1] majority logic

۳-۲- مدل بدون بازخورد و محدودیت‌های آن

در این قسمت محدودیت‌های مدل بدون بازخورد یا DAG یا چنددرختی[2] برای ساختار بررسی می‌شود. مطابق شکل (۴)، مدل چنددرختی، گرافی جهت‌دار است که در ساختار غیرجهت‌دار آن، بین هر دو گره، یا مسیری وجود ندارد یا یک مسیر وجود دارد و درنتیجه، گره‌ها و یال‌های قابل دسترسی از هر گره، یک درخت تشکیل می‌دهند [7].



(شکل-۴): مثالی از ساختار DAG
(Figure-4): Sample DAG Structure

پردازشِ گره ممکن است به شکل دقیق، احتمالاتی یا منطقی مدل‌سازی شود. در شکل (۴) بردار سیگنال‌های حاوی اطلاعات از طریق گره‌های ورودی (۱ تا ۴) به شبکه وارد شده، در چند لایه پردازش شده و در پایان، از طریق گره‌های خروجی (۵ تا ۷) خارج می‌شود.

نکته‌های زیر را در بارهٔ مدل DAG می‌توان مطرح کرد:

نکتهٔ ۱) شباهت عرضی افزایندهٔ سیگنال‌ها[3] (در پیشروی در شبکه): فرض کنید بردار ورودی X، به اندازهٔ آنتروپی شانون H(X) [31] اطلاعات در بر دارد. مدل DAG ساختار کانال تداخل[4] [31] را داشته و بنابراین نشت عرضی اطلاعات در آن رخ می‌دهد. اگر بردار X در عبور از شبکه، در لایهٔ m-ام به بردار Y_m و درنهایت به بردار خروجی Z تبدیل شود، داریم:

$$H(Z) \geq H(Y_{m+1}) \geq H(Y_m) \geq H(X) \qquad (۵)$$

یعنی آنتروپی بردار سیگنال بر اثر حرکت رو به جلو در شبکه و پردازش در لایه‌ها، افزایش می‌یابد.

نکتهٔ ۲) ضعف در پیروی از ورودی: به‌طور طبیعی به‌دلیل وجود حافظه در گره‌های پردازش‌گر، هر لایهٔ میانی، به مقدار مشخصی بر ویژگی متوسط‌گیری زمانی یا کندی عملکرد

[2] directed asyclic graph or multi-tree
[3] increasing transversal signal similarity
[4] interference channel

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]
[DOI: 10.29252/jsdp.17.2.112]

شبکه می‌افزاید؛ این موجب کاهش سرعت تغییر یا پهنای باند سیگنال در عمق شبکه می‌شود. این موضوع را با نامساوی پردازش داده‌ها[1] [31] نیز می‌توان توصیف کرد (هر پردازشی موجب از دست رفتن بخشی از اطلاعات می‌شود):

$$I(Z; X) \leq I(Y_{m+1}; X) \leq I(Y_m; X) \qquad (۶)$$

درجهٔ وفاداری خروجی نسبت به ورودی، به تعداد لایه‌های میانی بستگی دارد و با افزایش تعداد لایه‌ها کاهش می‌یابد.

نکتهٔ ۳) نبود بازخورد: به‌علت نبود بازخورد، پاسخ‌های شبکه در غیاب سیگنال‌های ورودی به‌سرعت از میان می‌روند.

ویژگی‌های شبکه‌های زیستی واقعی چنین است:

الف) بنا بر فرض، تعداد زیادی ژن (بیش از پنجاه) وجود دارد.

ب) تعداد کمی ورودی (کمتر از پنج)، سیگنال‌های ورودی را از شبکه‌های دیگر به GRN وارد می‌کنند [7].

پ) تعداد کمی خروجی (کمتر از پنج)، محصولات خروجی GRN را تشکیل می‌دهند [7].

ویژگی‌های الف تا پ بالا در مدل DAG، به معنای تعداد زیادی لایهٔ میانی و تعداد کمی گره در هر لایه است؛ بنابراین مدل شبکه شبیه به شکل (۴) با تعداد زیادی لایهٔ میانی است؛ زیرا در غیر این صورت باید تعداد لایه‌ها کم و تعداد گره در هر لایه زیاد باشد که این مشکل دیگری را ایجاد می‌کند: اگر تعداد کمی سیگنال ورودی، به تعداد زیادی سیگنال در لایه(های) میانی نگاشته شوند، از نظر پردازشی کار بیهوده‌ای انجام شده است؛ زیرا بُعد فضای سیگنال ورودی (حداکثر ۵) در لایه‌های میانی افزایش می‌یابد که نوعی اتلاف منابع در توصیف برداری با بُعد پایین در فضایی با بُعد بالاتر است.

نکته‌های سه‌گانهٔ بالا در کنار ویژگی‌های شبکه‌های واقعی، به نتایج زیر منجر می‌شود:

نتیجهٔ یک: نکتهٔ یک در کنار ویژگی‌های شبکه‌های واقعی به معنای شباهت زیاد مؤلفه‌های بردار خروجی شبکه است. این بدان معنا است که برای مقادیر مختلف بردارهای ورودی، مؤلفه‌های بردار خروجی به‌طورتقریبی یکسان هستند. این نتیجه از دیدگاه زیست‌شناسی صحیح نیست؛ زیرا شبکه‌های واقعی، در شرایط مختلف ورودی، محصولات متفاوتی تولید می‌کنند.

نتیجهٔ دو: نکتهٔ دو در کنار ویژگی‌های شبکهٔ DAG، به معنای کُندی شبکه در واکنش به تغییر ورودی‌ها است. با افزایش تعداد لایه‌های میانی، وفاداری در تعقیب تغییرات زمانی ورودی کمتر شده و بردار سیگنال خروجی کندتر از ورودی تحول می‌یابد. چنین نتیجه‌ای با واکنش سریع مشاهده شده در شبکه‌های واقعی زیستی هم‌خوانی ندارد.

نتیجهٔ سه: با توجه به نکتهٔ سه، با قطع سیگنال ورودی، خروجی‌های شبکه از میان می‌روند؛ اما در سیستم‌های واقعی، پاسخ‌ها تا مدتی ادامه می‌یابند؛ یعنی نوعی عملکرد خودمختار[2] در غیاب ورودی وجود دارد. دست‌یابی به چنین رفتاری، به وجود یک یا چند نوسان‌گر نیاز دارد و وجود نوسان‌گر، نیاز به بازخورد را مطرح می‌سازد که در مدل DAG جایی ندارد.

در مجموع و با توجه به مشخصه‌های شبکه‌های واقعی، مدل‌سازی شبکه‌های زیستی با مدل DAG قابل توجیه نیست. در مجموع، سیستمی با مدل DAG، بردار سیگنال ورودی را با کیفیتی نازل (از نظر سرعت و شباهت) تقلید می‌کند و توانایی عملکرد خودمختار نیز ندارد.

۳- تحلیل آزمایش‌های اختلال ژنی

مدل منطقی خانواده‌ای مهم از مدل‌های GRN است که امروزه مقبولیت یافته و شناخت ویژگی‌های آن اهمیت کاربردی فراوانی دارد. در مدل بولی، سیگنال خروجی گره متغیری باینری فرض می‌شود که دو وضعیت روشن (بیان‌شدن ژن) و خاموش (بیان‌نشدن) دارد و رابطه‌های منطقی بین فعالیت گره‌ها مطالعه می‌شود [7, 11, 20, 21, 29].

۳-۱- مدل بولی و شناسایی با آزمایش اختلال

در مدل بولی غیرتصادفی، تابع پردازش‌گر گره، تابعی بولی و غیرتصادفی است که یکی از دو مقدار صفر یا یک را (بسته به وضعیت ورودی‌های گره) اختیار می‌کند. شبکه‌ای با N گره، 2^N بردار حالت مجزا متناظر با بسط $[0, 1, \dots, 2^N - 1]$ دارد. در حالت عادی و در غیاب تغییر ورودی‌ها، شبکه در وضعیت مانا قرار دارد و بردار حالت (به‌طورتقریبی) ثابت است [7]. در صورت تغییر ورودی‌ها یا ایجاد اختلال در بخشی از شبکه، حالت گذرا ایجاد شده و پس از مدتی، بردار حالت به وضعیت (به‌طورتقریبی) ثابت مانا باز می‌گردد. جنبه‌های مختلفی از رفتار مدل‌های بولی، مانند پایداری و حالت گذرا موضوع پژوهش‌های متعددی بوده است [11, 20, 21, 29, 33].

شناسایی مدل بولی با استفاده از آزمایش‌های اختلال ژنی انجام می‌شود که در آن سیگنال خروجی گره(های)

[1] data processing inequality

[2] self regulation or autonomous reaction

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

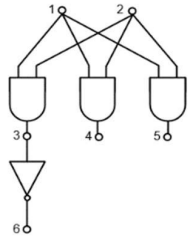
[DOI: 10.29252/jsdp.17.2.112]

شبکه تغییر داده می‌شود؛ به این صورت که سطح تولید محصول گره به‌صورت مصنوعی پایین آورده یا بالا برده شده و سپس به حال خود رها می‌شود [34] تا شبکه به حالت مانا برسد.

در آزمایش تک‌اختلال، هر بار"یکی" از گره‌ها دچار اختلال می‌شود. در پی ایجاد اختلال و پس از رسیدن شبکه به وضعیت مانا، ممکن است، سیگنال خروجی تعدادی از گره‌های دیگر نسبت به وضعیت قبل از ایجاد اختلال تغییر کند. چنین تغییر یا تغییرهایی، هم‌تغییری[1] نامیده می‌شود [19]. در آزمایش p-اختلال ($p > 1$) که نسبت به تک‌اختلال، پُرهزینه و زمان‌بر است [34]، سیگنال p گره تغییر داده می‌شود و سیگنال $N - p$ گره باقی‌مانده ممکن است تغییر کنند. تعداد کل آزمایش‌های p-اختلال برابر $\binom{N}{p}$ است.

هم‌تغییری، نشان‌دهندهٔ وجود رابطه(هایی) مستقیم یا غیرمستقیم بین گره‌های هم‌تغییر و گرهِ اختلال‌یافته است. از اطلاعات گره‌های هم‌تغییر در شناسایی ساختار شبکه استفاده می‌شود. برای آشنایی با مبانی شناسایی شبکه با آزمایش اختلال، مثالی از آزمایش تک‌اختلال آورده می‌شود.

مثال: آزمایش تک‌اختلال روی شبکهٔ بولی شکل (۵) انجام می‌شود [19]. در هر آزمایش، خروجی یکی از شش گره شبکه نسبت به حالت طبیعی آن معکوس شده و در پی آن، حالت مانای جدید شبکه محاسبه می‌شود. برای مثال گره یک که در حالت طبیعی در وضعیت روشن قرار دارد، خاموش می‌شود. ایجاد اختلال در گره یک، بر گره دو بی‌تأثیر است، اما موجب تغییر وضعیت گره‌های سه و چهار و پنج و شش می‌شود.



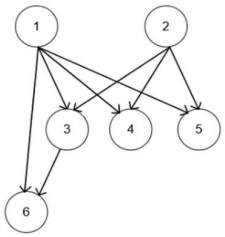
(شکل-۵): مثالی از شبکهٔ بولی [19]
(Figure-5): Example of a Boolean Network [19]

نتایج شش آزمایش در جدول (۲) ارائه شده است.

(جدول-۲): ماتریس تک-اختلال برای شبکهٔ بولی نمونه
(Table-2): Single-Perturbation Matrix for Example Network

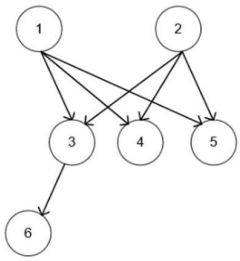
		g_1	g_2	g_3	g_4	g_5	g_6
	Wild type	**1**	**1**	**1**	**1**	**1**	**0**
1	g_1 off	0	1	0	0	0	1
2	g_2 off	1	0	0	0	0	1
3	g_3 off	1	1	0	1	1	1
4	g_4 off	1	1	1	0	1	0
5	g_5 off	1	1	1	1	0	0
6	g_6 on	1	1	1	1	1	1

مرحلهٔ بعدی در شناسایی مدل بولی آن است که نقشهٔ اولیه‌ای از ارتباطات علت و معلولی بین گره‌ها تهیه شود. شکل (۶) نسخهٔ اولیهٔ رابطهٔ بین ژن‌ها را نشان می‌دهد که بر اساس اطلاعات جدول (۲) تهیه شده است. در این شکل برای مثال، با توجه به هم‌تغییری ژن‌های سه و چهار و پنج و شش با اختلال ژن یک، یک یال ارتباطی بین ژن یک و هر یک از آنها قرار داده شده است.



(شکل-۶): نقشهٔ اولیه‌ای از ارتباطات علت و معلولی بین گره‌ها
(Figure-6): Preliminary Graph of the Causal Relations

استنتاج نقشهٔ تأثیرگذاری از نقشهٔ اولیه، مرحلهٔ بعدی است. در [19] از الگوریتم پوشش مجموعه[2] برای این منظور استفاده شده که اساس آن، جست‌وجوی کوچک‌ترین مجموعه‌هایی است که مجموعه‌های هم‌تغییری را پوشش دهند. نتیجه در شکل (۷) آورده شده است.



(شکل-۷): نقشهٔ تأثیرگذاری برای شبکهٔ نمونه
(Figure-7): Influence Map for Example Network

[1] co-variation

[2] set-covering

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]

پس از شناسایی ساختار شبکه، باید توابع بولی گره‌ها تعیین شود. این کار با جست‌وجو در جدول اختلال انجام می‌شود. برای تعیین رابطهٔ بین دو گره یک و دو با گره سه، ستون مربوط به این سه گره در جدول (۲) بررسی می‌شود که AND را نشان می‌دهد.

۲-۳- تعداد بهینهٔ اختلال برای شناسایی GRN

انجام آزمایش چنداختلال، نیازمند هزینه و زمان بسیار بیشتری در مقایسه با آزمایش تک‌اختلال است. نتایج آزمایش‌های چنداختلال به‌طورمعمول مفیدتر نیز در نظر گرفته می‌شود [35-37]؛ زیرا چنین به نظر می‌رسد که به‌علت ایجاد هم‌زمان چند اختلال، تعداد بیشتری هم‌تغییری در شبکه ایجاد شده و بنابراین اطلاعات بیشتری نیز حاصل می‌شود. در این قسمت، آزمایش‌های چنداختلال تحلیل شده و آزمایش بهینهٔ اختلال مشخص می‌شود. درواقع، پرسش زیر بررسی و پاسخ داده می‌شود:

آیا آزمایش‌های دو یا چنداختلال نسبت به آزمایش تک‌اختلال برتری دارند؟

پیش از قضیهٔ اصلی، لم تعادل (بین مجموع درجهٔ ورودی و خروجی گره‌های زیستی) بیان و اثبات می‌شود.

لم تعادل: اگر از گره‌های ورودی و خروجی شبکه صرف‌نظر شود (که تعداد آنها کم است)، هر ژن تنظیم‌شده توسط چند ژن، خود، تنظیم‌کنندهٔ چند ژن دیگر است؛ یعنی خروجی ژن به بیش از یک ژن دیگر متصل می‌شود.

اثبات: اگر لم تعادل در GRN درست نباشد، لم اولر یا لم دست‌دادن[1] نقض می‌شود که می‌گوید: در هر گراف دلخواه، حاصل‌جمع درجهٔ ورودی گره‌ها، $\sum \deg^-(v)$ با حاصل‌جمع درجهٔ خروجی گره‌ها، $\sum \deg^+(v)$ برابر است؛ یعنی:

$$\sum \deg^-(v) = \sum \deg^+(v) = |E| \qquad (۷)$$

$\deg(v)$ درجهٔ گره v و $|E|$ تعداد کل یال‌های شبکه است. بنابراین هر ژنِ تنظیم‌شده توسط K ژن، به‌طور متوسط K ژنِ دیگر را (از طریق محصول خود) تنظیم می‌کند.

بنا بر لم تعادل، درجهٔ خروجی برای "تمامی" گره‌های شبکه نمی‌تواند برابر یک باشد. درواقع درجهٔ خروجی گره، به‌طور متوسط، با درجهٔ ورودی گره برابر است.

قضیه (آزمایش بهینهٔ اختلال): آزمایش بهینه برای شناسایی زیرشبکه‌ها، آزمایش تک‌اختلال است.

توضیح: دو راهبرد برای شناسایی شبکه وجود دارد: ۱) شناسایی بیشترین تعداد ممکن از زیرشبکه‌ها (واحدهای سازنده)، ۲) شناسایی کل شبکه به‌صورت یک‌جا.

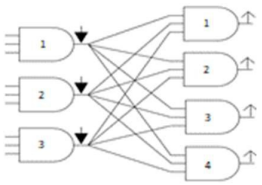
اثبات: به‌طور معادل سه گزارهٔ زیر اثبات می‌شود: ۱) احتمال شناسایی زیرشبکه‌ها با آزمایش‌های p-اختلال $(p > 1)$ پایین است و با افزایش p به‌سرعت کاهش می‌یابد، ۲) احتمال شناسایی یک‌جا با آزمایش‌های p-اختلال $(p > 1)$ بسیار پایین است، ۳) اطلاعات حاصل از آزمایش تک‌اختلال برای شناسایی صحیح تعدادی از زیرشبکه‌ها مفید است.

اثبات گزارهٔ یک: مجموعهٔ p-اختلال‌ها به دو گروه تقسیم می‌شود: مطلوب و نامطلوب. نمونهٔ p-اختلال مطلوب در شکل (۸) نشان داده شده است[2]. بر اساس لم تعادل، اگر درجهٔ ورودی گره برای مثال برابر سه باشد، آنگاه خروجی هر گره (در اینجا سه گره چپ)، به‌طور متوسط به ورودی سه گره دیگر (در اینجا سه گره سمت راست) اتصال می‌یابد. اعمال‌شدن کل p-اختلال‌ها به‌صورتی که در شکل (۸) نشان داده شده، مطلوب است، اما ممکن نیست؛ زیرا تعداد اختلال‌های مطلوب برابر با تعداد کل گره‌ها یعنی برابر N است؛ درحالی‌که تعداد کل p-اختلال‌ها یعنی $\binom{N}{p}$، با رشد p به‌سرعت افزایش می‌یابد و از آنجا که ساختار شبکه ناشناخته است، مشخص نیست که کدام اختلال‌ها به شکل مطلوب (یعنی به‌طوردقیق به ورودی‌های گره‌ها) اعمال می‌شوند. نسبت تعداد اختلال‌های مطلوب به کل اختلال‌ها، r_N^p برابر است با:

$$r_N^p = \frac{N}{\binom{N}{p}} \qquad (۸)$$

$$\text{for } N = 50, p = 2: \quad r_{50}^2 \approx 0.0$$
$$\text{for } N = 50, p = 3: \quad r_{50}^3 \approx 0.002$$

که درصدهای ناچیزی هستند. برای تک‌اختلال $r_N^1 = 1$، یعنی برای تمامی N، کل اختلال‌ها مطلوب به‌شمار می‌روند.



(شکل-۸): p-اختلال مطلوب (در اینجا p = 3)
(Figure-8): Proper Perturbation (here p = 3)

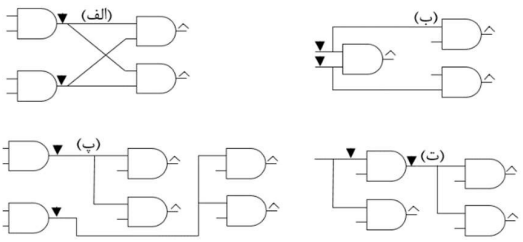
[1] Euler or handshaking lemma

[2] زیرا با توجه به حضور توابع راهبند (یا canalizing، یعنی AND و OR)، اعمال هم‌زمان چند اختلال به ورودی گره، احتمال تغییر خروجی گره را افزایش می‌دهد.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]
[DOI: 10.29252/jsdp.17.2.112]

یکی از حالت‌های نامطلوب آن است که دو اختلال در ورودی و خروجی یک گره قرار گیرند[1]. در این حالت، اطلاعات مفیدی به‌دست نمی‌آید؛ زیرا اختلال اعمال‌شده به ورودی گره نمی‌تواند تغییری را در خروجی که به‌طور اجباری تغییر یافته، ایجاد کند. برای وضوح بیشتر، نامطلوب‌بودن بیشینه p-اختلال‌ها برای $p = 2$ نشان داده می‌شود.

شکل (۹) چهار وضعیت ممکن را در اعمال دواختلال نشان می‌دهد. علامت ▼ مکان اعمال اختلال و علامت ۸ مکان مشاهدهٔ هم‌تغییری را نشان می‌دهد. بدون کاستن از کلیت تحلیل فرض می‌شود که توابع راه‌بند مشکلی ایجاد نمی‌کنند و اعمال اختلال به ورودی گره، خروجی آن را تغییر می‌دهد. فقط شکل (۹-الف) شبیه به شکل (۸) بوده و شرایط دواختلال مطلوب را دارد.



(شکل-۹): چهار وضعیت ممکن در دواختلال
(Figure-9): Four Possible Ways for 2-Perturbations

در شکل (۹-ب)، دو اختلال به دو ورودی گرهی اعمال شده که به خروجی‌های دو گره دیگر در سمت چپ‌شان متصلند (که در شکل نشان داده نشده‌اند) و بنابراین، حداکثر سه هم‌تغییری ایجاد می‌شود. این حالت، تا اندازه‌ای مفهوم دواختلال را در بر دارد، اما ابهام نیز دارد، زیرا مشخص نیست که کدام یک از این سه هم‌تغییری، با دواختلال اعمال‌شده به گره سمت چپ متناظر است.

در شکل (۹-پ)، دو اختلال به‌طور جداگانه به دو گره مختلف (هر اختلال به یک ورودی از یک گره) اعمال شده و حداکثر موجب چهار هم‌تغییری می‌شود. این حالت، درواقع دو تک‌اختلال همزمان است و به بروز ابهام بیشتری نیز منجر می‌شود؛ زیرا مشخص نیست که کدام یک از چهار هم‌تغییری، با کدام تک‌اختلال متناظر است.

شکل (۹-ت)، بیشترین ابهام را در اعمال دواختلال دارد، زیرا دو اختلال به‌طور هم‌زمان به ورودی و خروجی یک گره اعمال شده‌اند. در این حالت در بارهٔ عملکرد تک‌اختلالی نیز اطمینان وجود ندارد؛ زیرا مشخص نیست که کدام اختلال به ورودی و کدام یک به خروجی گره اعمال شده است. بنابراین در این وضعیت نیز هدف اصلی آزمایش دواختلال برآورده نمی‌شود.

بحث مثال بالا نشان می‌دهد که آزمایش p-اختلال فقط در وضعیت‌هایی مانند شکل (۹-الف) مفید است که p اختلال به‌شکل مطلوبی اعمال شده باشند. اما از آنجا که ساختار شبکه ناشناخته است، مشخص نیست که باید از کدام یک از وضعیت‌های شکل (۹) برای ساختن زیرشبکه‌ها استفاده شود؛ بنابراین احتمال شناسایی صحیح زیرشبکه‌ها به‌کمک آزمایش‌های p-اختلال ($p > 1$) پایین است و با توجه به رابطهٔ (۸)، با افزایش p، وضعیت به‌سرعت بدتر نیز می‌شود.

اثبات گزارهٔ دو: از آنجا که با افزایش p، تعداد شبکه‌های سازگار با آزمایش‌های p-اختلال کاهش می‌یابد[2]، ممکن است، چنین استدلال شود که با افزایش p، درنهایت به‌ازای مقداری مناسب از آن، شناسایی یک‌جا امکان‌پذیر می‌شود. این استدلال درصورتی درست است که اکثریت شبکه‌های سازگار با آزمایش‌های p-اختلال، از زیرشبکه‌هایی با ساختار درست تشکیل شده باشند تا انتخاب یکی از آنها به‌تصادف، با احتمال بالا، جواب درست را نتیجه دهد؛ اما چنین نیست؛ زیرا بر اساس استدلال مطرح‌شده در اثبات گزارهٔ (۱)، اکثریت شبکه‌های سازگار با آزمایش‌های p-اختلال ساختار درستی ندارند؛ زیرا از زیرشبکه‌هایی تشکیل شده‌اند که به‌احتمال بسیار بالا ساختاری غلط دارند. بنابراین احتمال شناسایی یک‌جا بر اساس آزمایش‌های p-اختلال بسیار پایین است.

اثبات گزارهٔ سه: در آزمایش تک‌اختلال، هر اختلال به‌احتمال بالا، به شناسایی تعدادی از ارتباط‌های مستقیم و غیرمستقیم کمک می‌کند و ابهامی در بارهٔ "وجود" این ارتباط‌ها وجود ندارد. درواقع هر آزمایش تک‌اختلال، یکی از سه نوع هم‌تغییری را ایجاد می‌کند:

۱- یک هم‌تغییری: که ارتباطی مستقیم را بین دو گره نشان می‌دهد که از "وجود" آن اطمینان وجود دارد.
۲- دو یا چند هم‌تغییری: که دست‌کم یکی از آنها ارتباط مستقیم است و بقیه ارتباط‌هایی غیرمستقیم هستند که به شناسایی ارتباط‌های مستقیم دیگری کمک می‌کنند.
۳- بدون هم‌تغییری: که به‌طورمعمول (و نه همواره) ورودی یا خروجی بودن آن گره را نشان می‌دهد.

[1] در این حالت، یکی از دو ژن به ورودی دیگری متصل است.

[2] زیرا برای p-ی بزرگتر، تعداد قیدها بیشتر است؛ برای مثال $\binom{N}{2}$ قید در دواختلال در مقایسه با N قید در تک‌اختلال.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]
[DOI: 10.29252/jsdp.17.2.112]

بنابراین از دو حالت نخست، همواره، و از حالت سوم، در بیشتر موارد، اطلاعاتی در بارهٔ ساختار زیرشبکه‌ها به‌دست می‌آید.

از دیدگاه شناسایی توابع بولی، تفاوتی بین آزمایش‌های تک‌اختلال و p-اختلال وجود ندارد؛ زیرا شناسایی توابع از طریق بررسی رابطهٔ بین ستون‌های متناظر با سیگنال‌ها در ماتریس اختلال مانند جدول (۲) امکان‌پذیر است.

این، اثباتِ بهینه‌بودن آزمایش تک‌اختلال از نظر شناسایی زیرشبکه‌ها را تکمیل می‌کند.

به‌طور کلی، شناسایی کامل و بی‌خطای شبکه با مدل بولی ناممکن است؛ زیرا حتی در آزمایش تک‌اختلال نیز (با توجه به قضیهٔ بالا) احتمال مشخصی برای خطا وجود دارد. استنتاج بخشی از زیرشبکه‌ها بر مبنای نتایج آزمایش تک‌اختلال امکان‌پذیر است و از آزمایش‌های p-اختلال می‌توان برای برطرف‌کردن بخشی از ابهام استفاده کرد [38].

۴- جمع‌بندی

در این مقاله، مسألهٔ مدل‌سازی GRN از دو دیدگاه مطالعه شد: ۱) محدودیت روش‌های کلاسیک مدل‌سازی GRN، ۲) نحوهٔ مدل‌سازی آزمایش‌های p-اختلال و دیدگاه‌های موجود.

انگیزهٔ اولیهٔ این پژوهش، پس از مشاهدهٔ مشکلی در زمینهٔ مدل‌سازی GRN ایجاد شد که عبارت است از استفادهٔ گسترده از مدل‌های نامناسبی که برخاسته از دیدگاه احتمالاتی نسبت به عملکرد GRN بوده است: مدل‌سازی احتمالاتی برای توصیف سازوکارهای مربوط به بیان ژن. مدل احتمالاتی محدودیت‌های مهمی را برای مدل‌سازی رفتار ژن‌ها ایجاد می‌کند و نتایج نادرستی را به‌دست می‌دهد که در بخش دوم مقاله بررسی شد.

در بخش دوم مقاله همچنین محدودیت‌های ناشی از کاربست مدل‌های بدون بازخورد در مدل‌سازی ساختار بررسی شد. نشان داده شد که این مدل که هم‌اکنون در مدل‌سازی شبکه‌های زیستی به‌کار می‌رود، محدودیت‌های مهمی داشته و توانایی مدل‌سازی برخی ویژگی‌ها و واقعیت‌ها را ندارد.

در بخش سوم و پایانی، مدل‌سازی بولی رفتار ژن و مسألهٔ کاربست آزمایش‌های اختلال برای شناسایی این مدل بررسی و نشان داده شد که اگر هدف، یافتن شبکه‌ای شامل زیرشبکه‌هایی با ساختار درست باشد، آزمایش تک‌اختلال، ابزار بهینه برای این منظور است و با کمترین هزینه، بیشترین اطلاعات را در بارهٔ زیرشبکه‌ها به‌دست می‌دهد.

یکی از یافته‌های مهم این مقاله آن است که نمی‌توان ساختار شبکه را - از صفر - بر اساس آزمایش‌های چنداختلال شناسایی کرد؛ زیرا همچنان که در قضیه اثبات شد، ابهام بسیار بالا است. بر اساس یافته‌های پژوهش، پیشنهاد می‌شود که نخست زیرشبکه‌های شبکهٔ جواب با آزمایش‌های تک‌اختلال شناسایی و سپس تعدادی آزمایش تکمیلی دو- یا چنداختلال برای برطرف‌کردن ابهام طراحی شود.

سپاس‌گزاری

از داوران محترم مقاله، برای دقت نظر و ارائهٔ پیشنهادهایی سازنده که برای بهبود ساختار و وضوح بیشتر در بیان مؤثر بود، سپاس‌گزاری می‌شود.

این پژوهش با حمایت کُرسی نظریهٔ اطلاعاتِ شبکه و امنیت، قرارداد شمارهٔ ۹۶/۵۳۹۷۵ (در ISSL) انجام شده است.

۵- مراجع 5-References

[1] G. Battail, "Heredity as an encoded communication process," *IEEE Trans. Information Theory*, vol. 56, no. 2, pp. 678-687, Feb. 2010.

[2] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 11, pp. 1021–1025, Nov. 1994.

[3] S. A. Salehi, et al, "Computing mathematical functions using DNA via fractional coding," *Nature Genetics*, May 2018.

[4] S. M. H. Tabatabaei Yazdi, et al, "Mutually uncorrelated primers for DNA-based data storage," *IEEE Trans. Information Theory*, vol. 64, no. 9, pp. 6283–6296, Sept. 2018.

[5] M. K. Gupta, "Quest for error correction in biology," *IEEE Eng. in Medicine and Biology Mag*, vol. 26, no. 1, Jan. 2006.

[6] B. Alberts, et al, *Molecular biology of the cell*, 6th edition, Garland Science, New York, 2014.

[7] S. Das, et al, *Handbook of research on computational methodologies in gene regulatory networks*, Hershey, New York, 2010.

[8] J. J. Pasternak, *An introduction to human molecular genetics: Mechanisms of inherited diseases*, 2nd edition, Wiley, New York, 2005.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]

[9] Beom S. Lee, et al, "A computational algorithm for personalized medicine in schizophrenia," *Schizophrenia Research*, vol. 192, pp. 131-136, Feb. 2018.

[10] N. Kornienko, et al, "Interfacing nature's catalytic machinery with synthetic materials for semi-artificial photosynthesis," *Nature Nanotechnology*, vol. 13, pp. 890–899, Oct. 2018.

[11] M.Banf, Seung Y. Rhee, "Computational inference of gene regulatory networks: Approaches, limitations and opportunities," *Biochimica et Biophysica Acta*, vol. 1860, no. 1, pp. 41-52, Jan 2017.

[12] N. Friedman, et al, "Using Bayesian networks to analyze expression data," *J. Computational Biology*, vol. 7, no. 3, pp. 601–620, March 2000.

[13] F. Fages, et al, "Influence networks compared with reaction networks: Semantics, expressivity and attractors," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 15, no. 4, pp. 1138 – 1151, July 2018.

[14] Y. Li, "The max-min high-order dynamic Bayesian network for learning gene regulatory networks with time-delayed regulations," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 13, no. 4, pp. 792-803, July 2016.

[15] H. Chen, et al, "Bayesian data fusion of gene expression and histone modification profiles for inference of gene regulatory network," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, doi: 10.1109/TCBB.2018.2869590, early access, Sep. 2018.

[16] M. Shi, et al, "Adaptive modelling of gene regulatory network using Bayesian information criterion-guided sparse regression approach," *IET Systems Biology*, vol. 10, no. 6, pp. 252-59, June 2016.

[17] S. Chan, et al, "Maximum a posteriori probability and time-varying approach for inferring gene regulatory networks from time course gene microarray data," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 12, no. 1, pp. 123 – 135, Jan. 2015.

[18] H. Xu, et al, "Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells," *PLoS One Computational Biology*, vol. 10, no. 8, pp. 1-14, Aug. 2014.

[19] S. Mehra, W. Hu, G. Karypis, "A Boolean algorithm for reconstructing the structure of regulatory networks," *Metabolic Engineering*, vol. 6, no. 4, pp. 326-39, Nov. 2004.

[20] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets, *J. of Theoretical Biology*, vol. 22, no. 3, pp. 437-67, March 1969.

[21] S. A. Kauffman, "The large-scale structure and dynamics of gene control circuits: an ensemble approach," *J. of Theoretical Biology*, vol. 44, no. 1, pp. 167-90, March 1974.

[22] J. I. Joo, et al, "Determining relative dynamic stability of cell states using Boolean network mode," *Nature Scientific Reports*, vol. 8, no. 1, online, Aug. 2018.

[23] www.humancellatlas.org, accessed Sep. 2018.

[24] H. P. Yockey, *Information theory, evolution and origin of life*, 2nd edition, Cambridge University Press, New York, 2011.

[25] S. L. Salzberg, et al, "Open questions: How many genes do we have?" *BMC Biology*, vol. 16, no. 1, online, Aug. 2018.

[26] Y. Lee, Qing Zhou, "Co-regulation in embryonic stem cells via context-dependent binding of transcription factors," *Bioinformatics*, vol. 29, no. 17, pp. 2162–68, Sept. 2013.

[27] A. J. M. Walhout, "What does biologically meaningful mean? A perspective on gene regulatory network validation," *Genome Biology*, vol. 12, no. 4, online, April 2011.

[28] M. Hecker, et al, "Gene regulatory network inference: Data integration in dynamic models - a review," *BioSystems*, vol. 96, no. 1, pp. 86-103, April 2009.

[29] W. Lee, W.S. Tzou, "Computational methods for discovering gene networks from expression data," *Briefings in Bioinformatics*, vol. 10, no. 4, pp. 408-23, July 2009.

[30] Q. Zhang et al, "Using single-index ODEs to study dynamic gene regulatory networks," *PLoS One Computational Biology*, vol. 13, no. 2, online, Feb. 2018.

[31] T. M. Cover, J. A. Thomas, *Elements of information theory*, 2nd edition, Wiley, New York, 2006.

[32] K. Do, P. Muller, M. Vannucci, *Bayesian inference for gene expression and proteomics*, Cambridge University Press, New York, 2006.

[33] P. Lin and S. P. Khatri, "Determining gene function in Boolean networks using Boolean satisfiability," *IEEE Int'l Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, San Antonio, Dec. 2012.

[34] T. E. Ideker, V. Thorsson, R. M. Karp, "Discovery of regulatory interactions through perturbation: Inference and experimental design," *Proc. Pacific Symposiums on Biocomputing*, pp. 305-316, Hawaii, Jan. 2000.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]

[35] A. Kaufman, M. Kupiec, E. Ruppin, "Multi-knockout genetic network analysis: the Rad6 example, *Proc. IEEE Computational Systems and Bioinformatics (CSB) Conference*, pp. 332–340, Stanford, California, Feb 2004.

[36] A. Kaufman, et al, "Quantitative analysis of genetic and neuronal multi-perturbation experiments," *PLoS One Computational Biology*, vol. 1, no. 6, online, Nov. 2005.

[37] R. Dehghannasiri, B. Yoon, Edward R. Dougherty, "Optimal experimental design for gene regulatory networks in the presence of uncertainty," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 12, no. 4, pp. 938-950, July 2015.

[38] A. R. Alizad-Rahvar, M. Sadeghi, "Integrative perturbation analysis of logic-based models of gene regulatory networks," *PLoS One* Computational Biology, Accepted, Oct. 2018.



رضا بیات دانش‌آموختهٔ مهندسی برق از دانشگاه‌های صنعتی خواجه نصیر، تربیت مدرس و یزد، و استادیار دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه سمنان است. زمینه‌های پژوهشی ایشان عبارت است از: نظریهٔ مخابرات، پردازش سیگنال، علم داده، بیوانفورماتیک.

نشانی رایانامهٔ ایشان عبارت است از:

Bayatr@semnan.ac.ir



مهدی صادقی دانش‌آموختهٔ بیوفیزیک از دانشگاه تهران و دانشیار پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری است. زمینه‌های پژوهشی ایشان عبارت است از: زیست‌شناسی نظری و محاسباتی، بیوانفورماتیک.

نشانی رایانامهٔ ایشان عبارت است از:

Sadeghi@nigeb.ac.ir



محمدرضا عارف دانش‌آموختهٔ مهندسی برق از دانشگاه تهران و استنفورد و استاد دانشکدهٔ مهندسی برق دانشگاه صنعتی شریف است. زمینه‌های پژوهشی ایشان عبارت است از: نظریهٔ اطلاعات، رمزنگاری.

نشانی رایانامهٔ ایشان عبارت است از:

Aref@sharif.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-14]

[DOI: 10.29252/jsdp.17.2.112]