Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال اول، شماره اول، بهار و تابستان ۱۴۰۲، صفحات ۱۰۷-۱۰۲

نقش گراف‌ها با ساختار درخت در برخی گرایش‌های علوم

محمد زینالی عظیم [۱*]، سعید علیخانی [۲]

[۱] مربی دانشکده مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد بستان‌آباد، بستان‌آباد، ایران

[۲] استاد دانشکده علوم ریاضی، دانشگاه یزد، یزد، ایران

چکیده

درختان، یکی از اساسی‌ترین کلاس‌ها در گراف‌ها هستند. آن‌ها نه‌تنها نقش کلیدی در نظریه گراف و ترکیبیات دارند، بلکه در بسیاری از زمینه‌های دیگر ریاضیات و همچنین در سایر علوم مانند زیست‌شناسی، شیمی و علوم کامپیوتر نیز ظاهر می‌شوند. این مقاله به بررسی خلاصه کاربردهای درختان در شیمی، زیست‌شناسی و کامپیوتر می‌پردازد.

مقاله پژوهشی

تاریخ دریافت:
۱۳۹۹/۱۲/۱۵

تاریخ پذیرش:
۱۴۰۰/۰۲/۱۸

کلیدواژه‌ها:
درخت، گراف، الگوریتم، علوم کامپیوتر

نویسنده مسئول:
mo.zeynali@gmail.com

doi : 10.22034/abmir.2022.2751

۱- مقدمه

درختان ساختار بسیار ساده‌ای دارند، بنابراین بسیاری از مسائل نظری در گراف که بسیار دشوار هستند، می‌توانند برای درختان بررسی و حل شود، و سؤالات الگوریتمی که ازنظر محاسبه برای گراف‌های عمومی سخت هستند، اغلب راه‌حل‌های کارآمد ساده‌ای برای درختان دارند. درعین‌حال، مسائل بسیار سختی نیز وجود دارد که به‌طور خاص به درختان مربوط می‌شود، به‌عنوان‌مثال حدس رینگل، که بیان می‌کند که گراف کامل $k_{(2n-1)}$ را می‌توان به $n-1$ کپی از درخت داده‌شده مرتبه n، به‌صورت یال-مجزا، تجزیه کرد. این حدس، اخیراً فقط برای n های بزرگ تأییدشده است [۱۰].

مثال دیگری از مسائل موجود در درختان، که به‌طور کامل باز است، حدس بازسازی درخت گراهام است: با شروع از یک درخت T گراف خطی $L(T)$ (گرافی که رئوس آن یال‌های T هستند و دو رأس مجاور هم هستند اگر و فقط اگر که یال‌های مربوطه در T مجاور باشند) و تکرارهای آن، یعنی $L(T), L(L(T)), L\left(L(L(T))\right),\ldots$ را در نظر بگیرید. حدس بازسازی درخت گراهام، بیان می‌کند که دنباله مراتب $|T|, |L(T)|, |L(L(T))|,\ldots$ به‌طور منحصربه‌فرد درخت T را مشخص می‌کند. چیز زیادی در مورد این حدس شناخته‌شده نیست. پیشرفت اخیر [٤] نشان می‌دهد که تعداد درختان مرتبه n که به‌طور منحصربه‌فرد مشخص می‌شوند، حداقل $e^{\Omega((\log n)^{\frac{3}{2}})}$ است، اما این هنوز با تعداد درختان مختلف که به‌صورت تصاعدی با n رشد می‌کنند، فاصله زیادی دارد. این مثال‌ها انواع مسائل جالبی که با درختان سروکار دارند، را نشان داده و همچنین نشان می‌دهند که درختان فقط ظاهراً ساده و قابل‌درک هستند. در ادامه به برخی از کاربردهای اساسی درختان در گرایش‌های مختلف می‌پردازیم [۳]

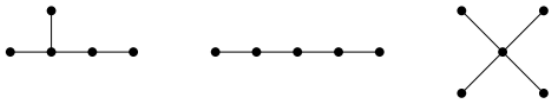
۲ – درختان فراگیر

یک درخت فراگیر گراف G، زیر گراف فراگیر G است که درخت نیز است. توجه کنید که درخت فراگیر یکتا نیست. به‌راحتی می‌توان دید که هر گراف همبند، دارای درخت فراگیر است. درختان فراگیر و تعداد آن‌ها از بحث‌های مهم در علومی مانند نانو و شیمی است. به روشی که به روش افزایشی معروف است، می‌توان درخت فراگیر یک گراف همبند را ساخت. در این روش یال‌هایی از گراف را به‌گونه‌ای برمی‌گزینیم که هیچ دوری ساخته نشود. این روند را ادامه می‌دهیم تا گراف القایی به‌وسیله این یال‌ها همه رئوس گراف را شامل شود. درروشی که به روش کاهشی مشهور است، هر دور گراف را انتخاب کرده و یکی از یال‌های آن را حذف می‌کنیم. این روند را ادامه می‌دهیم تا هیچ دوری در گراف باقی نماند. مساله مشهوری که به مساله کرتراند معروف است (۱۹۸۰ مطرح و سال ۱۹۸٦ به آن پاسخ داده شد) بیان می‌کند که اگر G یک گراف همبند ساده که شامل درخت‌های فراگیر با p و q برگ باشد و $p < r < q$ آنگاه G شامل درخت فراگیری با r برگ هم است. برای محاسبه تعداد درختان فراگیر رابطه بازگشتی وجود دارد، همچنین این پارامتر مهم را می‌توان با استفاده از قضیه درخت-ماتریس نیز محاسبه کرد. تعداد درختان فراگیر یک گراف می‌تواند عدد بزرگی باشد، برای مثال گراف پترسن دارای ۲۰۰۰ درخت فراگیر بر چسب‌دار است. منظور از درخت فراگیر مینیمم برای گراف همبند وزن‌دار، درختی است که بین درخت‌های فراگیر آن گراف، مجموع وزن یال‌های آن، کمترین مقدار ممکن باشد. در مسائلی که هدف ایجاد شبکه‌ای است که برای ایجاد ارتباط بین هر دو عضو آن هزینه‌ای باید بپردازیم و می‌خواهیم درنهایت در این شبکه بین هر دو عضو ارتباط وجود داشته باشد، درخت فراگیر کمینه همان کم‌هزینه‌ترین شبکه است. برای مثال فرض کنید در کشوری می‌خواهیم طوری جاده‌سازی کنیم که بتوان از هر شهری به هر شهر دیگری سفر کرد و هزینه ساخت جاده بین هر دو شهر رادایریم. این هزینه می‌تواند تابعی بر اساس فاصله دو شهر، آب‌وهوای بین دو شهر فاصله آن‌ها از شرکت راه‌سازی و ... باشد. برای پیدا کردن کم‌هزینه‌ترین راه، باید درخت فراگیر کمینه را بیابیم. که برای به دست آوردن آن از الگوریتم‌های مختلفی ازجمله الگوریتم کراسکال می‌توان استفاده کرد.

۳– درختان در شیمی

شیمی بر ریاضیات ترکیبانی برای مدت طولانی تأثیر گذاشته است، یک مثال ابتدایی، شمارش ترکیبات شیمیایی، که انگیزه‌ای آشکار در کار پیشگامانه پولیا [۱۱] در مورد آنچه اکنون به‌عنوان روش شمارش پولیا شناخته‌شده است، است.

درختان، نقش برجسته‌ای در این زمینه به‌عنوان مدل‌هایی برای مولکول‌های بدون دور (غیرحلقوی) دارند. به‌عنوان‌مثال، سه ایزومر اسکلتی پنتان C_5H_{12} ،که مطابق باکلاس‌های ایزومورفیسم درختان مرتبه پنج است، همان طور که در شکل ۱ نشان داده‌شده است، وجود دارد.



شکل (۱): درختان مرتبه ۵

زمینه نظریه گراف شیمیایی در سال‌های اخیر رشد زیادی داشته است که از دلایل آن می‌توان به مشمول شدن دو کد MSC جدید در نسخه ۲۰۲۰ اشاره کرد: گراف‌های شیمیایی و شاخص‌های گرافیکی $05C_{92}$ شاخص‌های گرافیکی (همچنین پس از ارائه شاخص هوسویا، شاخص‌های توپولوژیکی نامیده می‌شود) متغیرهایی هستند (معمولاً به‌صورت اعداد صحیح و یا حقیقی) که متناظر گراف‌ها هستند. در خیلی از موارد مشاهده‌شده است که این شاخص‌ها با خواص فیزیکی و شیمیایی مولکول‌های متناظر، هماهنگی و ارتباط خوبی دارند. یک نمونه برجسته، شاخص وینر یک گراف است، که به‌عنوان مجموع فواصل بین همه جفت رئوس تعریف می‌شود.

این در سال ۱۹۴۷ توسط هری وینر[1] در مقاله [۱۳] به‌عنوان شماره مسیر در ارتباط با نقطه‌جوش پارافین معرفی شد. همان‌طور که وینر درگذر زمان متوجه می‌شود، آن را می‌توان برای درختان به‌صورت زیر محاسبه کرد: برای هر یال e حاصل‌ضرب اندازه‌های دو مؤلفه همبندی به وجود آمده از حذف یال e، را در نظر بگیرید. سپس تمام جواب‌های حاصل را جمع کنید. به‌راحتی می‌توان فهمید که این حاصل‌ضرب، تعداد دفعاتی را نشان می‌دهد که یال e در یک (کوتاه‌ترین) مسیر بین دو رأس پدیدار می‌شود، بنابراین این مجموع درواقع برابر با شاخص وینر است. این ویژگی ساده و درعین‌حال شیک و مفیدِ شاخص وینر، یک مثال اولیه از یک قضیه در نظریه گراف شیمیایی است که درواقع در یک مجله شیمی ظاهر شد. جنبه‌های مختلفی برای نظریه گراف شیمیایی فعلی وجود دارد، اما معمولاً محققان به کران‌های شاخص‌های مختلف مرتبط با گراف‌ها، روابط بین شاخص‌های مختلف و سؤالات حدی که حداکثر یا حداقل یک شاخص را برای یک خانواده خاصی از گراف‌ها (به‌عنوان‌مثال، همه درختان n رأسی پیدا می‌کنند، علاقه‌مند هستند. درختان به‌طور سنتی نقش برجسته‌ای در این زمینه‌دارند، مانند انواع مختلف گراف‌های شبه–درخت که در شیمی مرتبط هستند، به‌عنوان‌مثال، گراف‌های با تعداد محدود دور یا دورهایی با طول‌های خاص مانند شش‌ضلعی. ساختارهای خاصی به‌طور مکرر در این زمینه رخ می‌دهند. بدون شک، به‌عنوان‌مثال، مسیر و ستاره معمولاً درختانی هستند که حداقل و حداکثر یک شاخص گرافیکی را به دست می‌آورند. مثال دیگر، درختان حریصانه است که مکرراً زمانی رخ می‌دهد که محدودیت‌های درجه تحمیل می‌شود. برای مثال بخش‌های ۲.۶.۲ و ۳.۲.۱ از [۱۲] را ببینید.

٤ – درختان در زیست‌شناسی

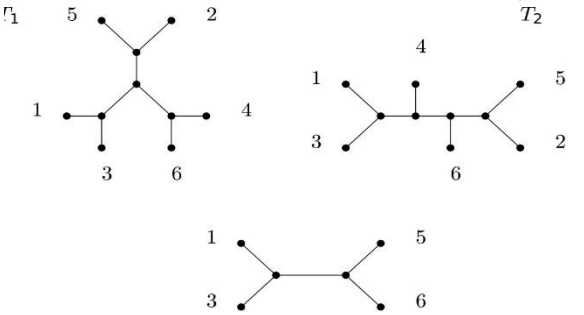
در زیست‌شناسی، درختان به‌طور برجسته به‌عنوان درختان فیلوژنتیک وجود دارند که روابط تکاملی بین گونه‌های مختلف را نشان می‌دهد. گونه[2]، متناظر به برگ درخت است، و ریشه (اگر وجود داشته باشد) به رایج‌ترین مسائل جدِمشترک[3] متناظر است. مسائل ریاضی مربوط به درختان فیلوژنتیک، اغلب با برنامه‌های محاسباتی انگیزه می‌گیرند.

یک مثال معمولی برای نشان دادن نوع سوالاتی که در این زمینه وجود دارد، به زیردرخت‌های توافق ماکزیمم[4] مربوط می‌شود. در این‌جا، درختان باینری با برگ‌های بر چسب‌دار را در نظر می‌گیریم، که در آن برگ‌ها دارای برچسب {1 و 2 و 3 و...n} هستند

[1] Harry Wiener
[2] species
[3] Common ancestor
[4] Maximum agreement subtrees

(می‌توان آن‌ها را نماینده گونه‌های مختلف تصور کرد). درخت باینری را می‌توان به زیرمجموعه S از برگ‌ها به روش طبیعی محدود کرد: کوچک‌ترین زیر درختی که همه این برگ‌ها را در خود دارد را در نظر گرفته، سپس، تمام رئوس درجه‌دو را سرکوب می‌کند (حذف رأس درجه‌دو و افزودن یک یال بین دو همسایه آن) تا یک درخت باینری جدید به دست آورد.

دو درخت مرتبه n با برگ‌های برچسب‌گذاری‌شده T_1 و T_2 را روی زیرمجموعه S از $\{1, 2, 3, \ldots, \}$ موافق گوییم، اگر محدودیت‌های مربوط به S یکسان باشد (یعنی برچسبی وجود دارد که ایزومورفیسم را حفظ می‌کند). به‌عنوان‌مثال، دو درخت در شکل ۲ روی مجموعه $\{1, 3, 5, 6\}$ موافق هستند. حداکثر اندازه مجموعه S که درختان T_1 و T_2 روی آن موافق هستند، با $MST(T_1,T_2)$ نشان داده می‌شود. این کمیت شبیه به طولانی‌ترین زیردنباله مشترک دو رشته یا جایگشت است. تا همین اواخر، این‌یک سؤال باز بود که این کمیت $ST(T_1,T_2)$ چقدر می‌تواند ازنظر تعداد برگ کوچک باشد. مارکین در مقاله اخیر خود [۷]، با بهبود کران پایین قبلی که از مرتبه $\sqrt{\log n}$ بود [۸]، ثابت کرد که می‌نیمم از مرتبه $\Theta(\log n)$ است.



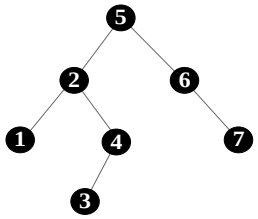
شکل (۲): دو درخت T_1 و T_2 به همراه تحدیدشان $\{۶$ و ۵ و ۳ و $۱\}$

برای جفت درختان تصادفی، مرتبه بزرگی اندازه متوسط یک زیر درخت توافق ماکزیمم، هنوز مشخص نیست. کران‌های بالای $O(\sqrt{n})$ برای هر دو مدل استاندارد فیلوژنتیک، یعنی مدل یکنواخت و مدل یول هاردینگ برقرار است. از سوی دیگر، برنشتاین، هو، لانگ، استیل، سنت جان، و سالیوانت [۱] کران‌های پایین $\Omega(n^{1/8})$ (مدل یکنواخت) و $\Omega(n^{0.344})$ (یول–هاردینگ) را ارائه کردند. شبیه‌سازی‌ها نشان می‌دهد که توان واقعی نزدیک به $1/2$ است، [۲] را ببینید. برای درختان تصادفی هم‌شکل (اما با برچسب‌های متفاوت)، اخیراً نشان داده‌شده است که ترتیب اندازه مورد انتظار $\Theta(\sqrt{n})$ است [۹].

۵ – درختان در علوم کامپیوتر

درختان به‌عنوان یک ساختار داده مفید هستند و اساس آن را بسیاری از الگوریتم‌ها تشکیل می‌دهند، برای مثال بخش ۲.۳ از [۶] را ببینید. به‌عنوان‌مثال، درختان، داده‌های سلسله مراتبی را ذخیره می‌کنند و می‌توان از آن‌ها برای جستجو و مرتب‌سازی استفاده کرد. درخت‌های جستجوی دودویی، نمونه‌ای معمولی از یک ساختار داده درختی هستند و همچنین به الگوریتم معروف مرتب‌سازی سریع مرتبط هستند.

در جستجوی دودویی درخت، برچسب‌ها در شاخه سمتِ چپِ ریشه، همه کوچک‌تر از برچسب‌ها در شاخه سمتِ راستِ ریشه هستند. همین امر در مورد دوشاخه دیگر در بقیه رئوس صدق می‌کند. شکل ۳ را ببینید.



شکل (۳): درخت جستجوی باینری

درختان برای مدت طولانی درزمینه تجزیه‌وتحلیل بدترین حالت و حالت متوسط الگوریتم‌ها موردمطالعه قرارگرفته‌اند؛ برای مثال فصل ۶ مرجع [۱۲] را ببینید. پارامترهایی مانند ارتفاع (بیشترین فاصله ریشه تا برگ) معیارهایی برای عملکرد این الگوریتم‌ها در این زمینه هستند. مسائل شمارشی مربوط به درخت‌ها و ساختارهای شبه–درخت و همچنین ویژگی‌های درختان تصادفی اغلب موردتوجه است، به‌خصوص زمانی که عملکرد متوسط الگوریتم‌ها در نظر گرفته شود. راه‌های مختلفی برای تولید درخت تصادفی وجود دارد، و خواص ساختاری، اغلب بسته به مدل خاص تصادفی استفاده‌شده، متفاوت است. شاید ساده‌ترین و راحت‌ترین راه برای تولید درخت تصادفی، مدل‌های یکنواختی باشد که در

آن‌یک درخت به‌طور تصادفی از بین همه درختان دارای n رأس انتخاب می‌شود.

یک روش کلاسیک برای تولید درختان با یک فرآیند تصادفی، درختان گالتون–واتسون است: این درختان با یک ریشه و یک نوع توزیع فرزندان، شروع می‌کنند. در مرحله k- ام، تمام رئوس از نسل k به‌طور مستقل بر اساس توزیع، فرزندان را تولید می‌کند.

درختان گالتون–واتسون و درختان یکنواخت تصادفی، مدل‌هایی هستند که در آن درختان معمولی دارای مرتبه n و ارتفاعی از مرتبه $\sqrt{n}$ هستند.

مدل‌های دیگر درختانی را تولید می‌کنند که ارتفاع آن‌ها معمولاً فقط از مرتبه لگاریتمی است: مثال‌ها شامل درختان جستجوی دودویی (که مدل احتمالی آن‌ها درواقع معادل مدل یول–هاردینگ ذکرشده در فیلوژنتیک است) و درختان بازگشتی، جایی که با یک رأس تنها شروع می‌شود و n- امین رأس به یکی از n-1 راس‌ها به‌طور یکنواخت تصادفی متصل است. ارتفاع، تنها یکی از بسیاری از خواص ساختاری درختان است که برای بسیاری از انواع مختلف درختان تصادفی موردمطالعه قرارگرفته است. نتایج بسیار دقیق، نه‌تنها در اندازه بزرگی، بلکه در توزیع، اغلب در [٤] در دسترس هستند.

٦ – درخت تصمیم

درخت تصمیم یک ابزار پشتیبانی تصمیم است که از یک مدل درخت مانند از تصمیمات و پیامدهای احتمالی آن‌ها، ازجمله نتایج رویدادهای شانسی، هزینه‌های منابع و ابزار استفاده می‌کند. این‌یکی از راه‌های نمایش الگوریتمی است که فقط شامل دستورات کنترل شرطی است. درخت‌های تصمیم معمولاً در تحقیق عملیات، به‌ویژه در تحلیل تصمیم‌گیری، برای کمک به شناسایی استراتژی‌هایی که به‌احتمال‌زیاد به یک هدف می‌رسند، استفاده می‌شوند، اما همچنین ابزاری محبوب در یادگیری ماشین هستند.

درخت تصمیم یک ساختار فلوچارت مانند است که در آن هر گره داخلی یک «آزمون» را بر روی یک ویژگی نشان می‌دهد (به‌عنوان‌مثال این‌که آیا یک سکه به سمت رو می‌آید یا پشت). هر شاخه نشان‌دهنده نتیجه آزمایش است، و هر برگ نشان‌دهنده یک ویژگی است. مسیرهای ریشه تا برگ نشان‌دهنده قوانین طبقه‌بندی هستند. در تجزیه‌وتحلیل تصمیم، یک درختِ تصمیم و نمودار تأثیر نزدیک مرتبط، به‌عنوان یک ابزار پشتیبانی تصمیم بصری و تحلیلی استفاده می‌شود، جایی که مقادیر مورد انتظار (یا مطلوبیت مورد انتظار) جایگزین‌های رقیب محاسبه می‌شود.

برای تعریف ساده الگوریتم درخت تصمیم در یک جمله می‌توان گفت، این الگوریتم شبیه بازی ۲۰ سوالی است. بدین‌صورت که در طول حل مساله، یک گراف درخت رسم می‌شود که گره‌های میانی آن سوال‌ها هستند و گره‌های برگ آن خروجی طبقه‌بندی را مشخص می‌کنند. یک گراف ساده را که یک الگوریتم درخت تصمیم را نشان می‌دهد بررسی می‌شود.

۷– نتیجه‌گیری

هدف این مقاله کوتاه این است که نشان دهد درختان و ساختارهای درختی در چه متن‌ها و موضوعات متفاوتی ظاهر می‌شوند و چه سؤالات ریاضی پیرامون درختان در زمینه‌های مختلف موردتوجه است.

References

[1] Daniel Irving Bernstein, Lam Si Tung Ho, Colby Long, Mike Steel, Katherine St. John, and Seth Sullivant, Bounds on the expected size of the maximum agreement subtree, SIAM J. Discrete Math. 29, 2015, no. 4, 2065–2074, DOI 10.1137/140997750. MR3416130

[2] Daniel Irving Bernstein, Lam Si Tung Ho, Colby Long, Mike Steel, Katherine St. John, and Seth Sullivant, Bounds on the expected size of the maximum agreement subtree, SIAM J. Discrete Math. 29, 2015, no. 4, 2065–2074, DOI 10.1137/140997750. MR3416130

[3] Joshua Cooper, Bill Kay, and Anton Swifton, Graham’s tree reconstruction conjecture and a Waring-type problem on partitions, J. Comb. 9, 2018, no. 3, 469–488, DOI 10.4310/JOC.2018.v9.n3.a3. MR3809644

[4] Michael Drmota, Random trees: An interplay between combinatorics and probability, SpringerWienNewYork, Vienna, 2009, DOI 10.1007/978-3-211-75357-6. MR2484382

[5] Donald E. Knuth, The art of computer programming. Vol. 1: Fundamental algorithms,

Addison-Wesley, Reading, MA, 1997. Third edition [of MR0286317]. MR3077152

[6] Alexey Markin, On the extremal maximum agreement subtree problem, Discrete Appl. Math. 285, 2020, 612–620, DOI 10.1016/j.dam.2020.07.007. MR4124794

[7] Daniel M. Martin and Bhalchandra D. Thatte, The maximum agreement subtree problem, Discrete Appl. Math. 161, 2013, no. 13-14, 1805–1817, DOI 10.1016/j.dam.2013.02.037. MR3056987

[8] Pratik Misra and Seth Sullivant, Bounds on the expected size of the maximum agreement subtree for a given tree shape, SIAM J. Discrete Math. 33, 2019, no. 4, 2316–2325, DOI 10.1137/18M1213695. MR4036089

[9] R. Montgomery, A. Pokrovskiy, and B. Sudakov, A proof of Ringel's conjecture, Geom. Funct. Anal. 31, 2021, no. 3, 663–720, DOI 10.1007/s00039-021-00576-2. MR4311581

[10] G. P\' olya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen (German), Acta Math. 68, 1937, no. 1, 145–254, DOI 10.1007/BF02546665. MR1577579

[11] Robert Sedgewick and Philippe Flajolet, An introduction to the analysis of algorithms, Addison-Wesley-Longman, 1996.

[12] Stephan Wagner and Hua Wang, Introduction to chemical graph theory, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2019. MR3837106

[13] H. Wiener, Structural determination of paraffin boiling points, J. Am. Chem. Soc. 69, 1947, 17–20.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 1, No. 1, Spring & Summer 2023, 102-107

The role of graphs with tree structure in some sciences

Mohammad Zeynali Azim[1*] Saeid Alikhani[2]

[1] Computer Engineering Department, BostanAbad Islamic Azad University, BostanAbad , Iran

[2] Department of Mathematical Sciences, Yazd University, Yazd, Iran

Article Information

Original Research Paper

Received:
27 June 2022

Accepted:
21 September 2022

Keywords:
Tree, Chemistry, Computer, Decision

Corresponding Author*:
mo.zeynali@gmail.com

Abstract

Trees are one of the most basic classes in graphs. They not only play a key role in the theory of graphs and combinatorics, but also in many other fields of mathematics, as well as in other sciences such as biology, chemistry, and computer science. In this article, we review the applications of trees in chemistry, biology, and computers.

: 10.22034/abmir.2022.2751