بسم الله الرحمن الرحيم

دانشگاه قم

دانشکده علوم پایه

پایان نامه دوره کارشناسی ارشد ریاضی کاربردی

عنوان:

توده‌ها در ساختارهای کانونی شبه گره RNA

استاد راهنما:

دکتر غلامحسن شیردل

استاد مشاور:

دکتر علی اصغر فروغی

نگارنده:

یاور نجات پور شرف آباد سفلی

تابستان ۱۳۸۹

تقدیم به:

مادر فرشته خصالم و پدرم که فروغ مهرشان صیقل روحم است

و توصیفشان را سینه‌ای به وسعت دریا باید تا بتوان اندکی از مهربانی

بسیارشان را در کلام وصف آورد.

تشکر و قدردانی

حمد و سپاس خدای را که توفیق کسب دانش و معرفت را به ما عطا فرمود. در اینجا بر خود لازم می دانم از تمامی اساتید بزرگوار، به ویژه اساتید دوره کارشناسی ارشد که در طول سالیان گذشته مرا در تحصیل علم و معرفت و فضائل اخلاقی یاری نموده اند تقدیر و تشکر نمایم.

از استاد گرامی و بزرگوار جناب آقای دکتر شیردل که راهنمایی اینجانب را در انجام تحقیق، پژوهش و نگارش این پایان نامه تقبل نموده اند نهایت تشکر و سپاسگزاری را دارم.

از جناب آقای دکتر فروغی به عنوان استاد مشاور که با راهنمایی خود مرا مورد لطف قرار داده‌اند کمال تشکر را دارم.

از اساتید گرامی و بزرگوار جناب آقای دکتر فرجامی و آقای دکتر احمدی نیا که نظارت پایان نامه مرا بر عهده داشتند، کمال تشکر را دارم.

چکیده

ساختارهای RNA بصورت پیکره حلزونی شکل از دنباله های RNA می باشد که به وسیله زنجیره هایی از نوکلئوتیدهای آدنین، گوانین، سیتوزین و اوراسیل توسط قوانین جفت پایه ای واتسن و کریک به صورت (A-U , U-G) و (G-C) پیوند هیدروژنی برقرار می کنند.

پیش بینی ساختار RNA یکی از مسائل مهم در علم بیوانفورماتیک می باشد، که توجه ویژه‌ای در سال های اخیر به ان شده است. در این پایان نامه روش های شمارش تعداد ساختارهایRNA، هم چنین توزیع توده ها در ساختار های RNA مورد بررسی قرار گرفته است. یک ساختار RNA را k-نامتقاطع گویند، هرگاه یک مجموعه k تایی از کمان هایی را که دو به دو متقاطع اند را شامل نمی شوند و یال های متقاطع گذرنده دوطرفه بیشتر از k-1 نباشند. و□-کانونی گویند، هر گاه یال های مشمول در یک توده حداقل τباشند.

واژه های کلیدی : ساختار RNA، گراف جهت دار k-نامتقاطع، ساختار کانونی RNA، توده.

فهرست مطالب

عنوان | **صفحه**

فهرست جداول

عنوان **صفحه**

.

فهرست شکل ها و نمودارها

شکل(۴-۳). نگاشت $c:T_{k,\tau}(n,t)\rightarrow\bigcup_{t\tau\leq h\leq\left[\frac{n}{۲}\right]}C_k^*(n-۲(h-t),t)$ ۸۲

که در دو گام بدست آمده است. اول انقباض توده‌ها و ثانیاً بر چسب

مجدد دیاگرام حاصل.

نمودار(۴-۱). توزیع حد مرکزی مربوط به تعداد توده‌ها در ساختارهای ۷۶

k- نامتقاطع و τ- کانونی، RNA.

فهرست علائم

علامت جایگشت α	$sign\ (\alpha)$
حاصل ضرب نیمه مستقیم H توسط K	H
دترمینان ماتریس A	$det\ A$
مجموعه $\{1, \dots , n\}$	$[n]$
f و g به طور مجانبی هم ارزند	$f(n) \sim g(n)$
f، اوی بزرگ g است	$f(z)=O(g(z))$
f، اوی کوچک g است	$f(z)=o(g(z))$
تابع بسل نوع مرتبه اول و از مرتبه α	$I_\alpha(z)$
تابع گاما	$\Gamma(x)$
ضریب z^n در بست مک لورن تابع $f(z)$	$[z^n]\ f(z)$
n امین عدد کاتالان	C_n
e^x	$exp(x)$

مقدمه

در این پایان نامه یک تقریب مناسب برای تعداد ساختارهای شبه گره k-نامتقاطع[1] RNA [2] و همچنین توزیع توده‌ها در ساختارهای کانونی RNA ارائه می‌شود.

این مبحث در حوزه ریاضیات زیستی[3] قرار دارد که اخیرا بسیار مورد توجه قرار گرفته و تحقیقات زیادی در حوزه‌های مختلف آن انجام می‌شود. ریاضیات زیستی یک حوزه بین رشته‌ای از مطالعات اکادمیک است که فرایند‌های زیستی را با استفاده از ابزارها و روشهای ریاضی مدل‌سازی می‌کند.

این رشته (بیومتمتیک) هم کاربرد تئوریک و هم کاربرد عملی در تحقیقات بیولوژیکی دارد. بعلت تنوع زیاد دانش‌های تخصصی، تحقیقات ریاضیات زیستی اغلب با همکاری بین ریاضی دانان، فیزیک دانان، زیست شناسان، جانور شناسان و شیمیدانان انجام می گیرد.

DNA[4] برای دهه‌های متمادی ستاره آسمان زیست شناسی مولکولی بوده است. اما با درک اهمیت و توانایی RNA که برای مدتی مدید فقط به عنوان نسخه بردار ژن‌های موجود در مارپیچ دوگانه شناخته می‌شد، می‌بایست صحنه را روز به روز با RNA به اشتراک گذارد. از نقطه نظر *RNA*، *DNA* تنها یک بایگانی غیر فعال است، یک دفترچه تلفن کند ذهن، حال

[1] *K-noncrossing*
[2] *Ribo Nucleic Acid*
[3] *Biomathematics*
[4] *Deoxyribonucleic Acid*

آنکه این RNA است که شماره‌ها را بازیابی می‌کند، ارتباطات را برقرار می‌سازد و تعیین می‌کند که هر ارتباط چه مدت بطول می‌انجامد. اوری مک لود[۱] و مکارتی[۲] نخستین بار در سال ۱۹۴۴ ثابت کردند که DNA حاوی اطلاعات ژنتیکی است. از آنجا که اطلاعات در ترتیب واحدهای منوپلیمرها نهفته است، باید مکانیسمی برای تکثیر یا همانند سازی این اطلاعات خاص وجود داشته باشد که دقت زیادی را داشته باشد. مطالعه این شاخه از ریاضیات نیاز به مفاهیم پایه‌ای گسترده‌ای در ریاضیات دارد. به عبارت دیگر مفاهیمی در نظریه گراف، ترکیبات، توابع مختلط و آمار احتمال مورد استفاده قرار می گیرد.

به همین دلیل در فصل اول، برخی از این مفاهیم و تعاریف آورده شده است.

در فصل دوم ابتدا مفاهیم زیست شناسی، و برخی مطالب ترکیبی ترسیم از قبیل تابلوهای یانگ[۳]، الگوریتم RSK [۴] و حجره وایل [5] را بیان می‌کنیم.

در فصل سوم با استفاده از مطالب مطرح شده در فصل دوم تابع مولد انطباق‌های k- نامتقاطع را محاسبه کرده و آنالیز تقریبی از تطابق‌های k- نامتقاطع را بدست می‌آوریم.

و در نهایت مبحث اصلی این پایان نامه قضایا و نتایجی است که در فصل چهارم بدست می‌آیند و شامل توزیع توده‌ها در ساختارهای کانونی RNA می باشند.

---

[1] Mack lod.
[2] Mackarti
[3] Young
[4] Robinson-schented-knuth
[5] Weyl

فصل اول

تعاریف و مقدمات

۱-۱ تعاریف گراف و مفاهیم وابسته به آن

تعریف (۱-۱-۱): یک گراف ساده G با n رأس و m یال متشکل از مجموعهٔ رأسهای $V(G)=\{v_1,\dots,v_n\}$ و مجموعه یالهای $E(G)=\{e_1,\dots,e_n\}$است، که در آن هر یال، یک جفت نامرتب از رأسهاست. به جای $\{u,v\}$ می‌نویسیم uv. اگر $uv\in E(G)$ آنگاه u و v مجاور هستند. این مطلب را با $u \leftrightarrow v$ نشان می‌دهیم، یعنی «u مجاور v است». رأسهای مشمول در یک یال e نقاط پایانی آن هستند.

تعریف (۱-۱-۲): یک گراف جهت‌دار ساده یا دیگراف[1] سادهٔ G متشکل از یک مجموعه رأسهای $V(G)$ و یک مجموعه یالهای $E(G)$ است، که در آن هر یال یک جفت مرتب از رأسهاست. به جای یال (u,v) می‌نویسیم uv، که u دُم و v سر آن است. وقتی $uv\in E(G)$، می‌نویسیم $u\to v$، یعنی «یالی از u به v وجود دارد».

تعریف (۱-۱-۳): مکمل یک گراف ساده G را با $\overline{G}$ نشان می‌دهیم و گرافی است با همان مجموعه رأسهای G، به طوری که u و v در $\overline{G}$ مجاورند اگر و فقط اگر u و v در G مجاور نباشند.

[1] *Digraph*

تعریف(۱-۱-۴): یک زیرگراف از یک گراف G، یک گراف H است به طوری که $V(H) \subseteq V(G)$ و $E(H) \subseteq E(G)$، که به صورت $H \subseteq G$ می‌نویسیم و می‌گوییم «G شامل H است».

تعریف (۱-۱-۵): یک زیرگراف القایی از G زیرگرافی مانند H است به طوری که هر یال G مشمول در $V(H)$ متعلق به $E(H)$ باشد.

اگر H یک زیرگراف القایی از G با مجموعه رأسهای S باشد آنگاه می‌نویسیم $H = G[S]$ و می‌گوییم H زیرگراف G «القاء شده به وسیله S» است.

تعریف (۱-۱-۶):: اگر H زیرگراف G باشد و $|V(H)| = |V(G)|$ در این صورت H را زیرگراف پوشا (فراگیر) G می‌نامیم.

تعریف (۱-۱-۷): گراف ساده‌ای که در آن هر دو راس با هم همجوار باشند را گراف کامل می‌نامیم. گراف کامل با n راس با K_n نمایش می‌دهیم.

تعریف (۱-۱-۸): گراف ساده‌ای که در آن بتوان مجموعه راس‌ها را به دو بخش چنان افراز کنیم که هیچ دو عضوی از یک بخش با هم همجوار نباشند را یک گراف دوبخشی گوییم.

تعریف (۱-۱-۹): گراف دوبخشی را کامل نامیم هرگاه یک راس از یک بخش با تمام راس‌های بخش دیگر همجوار باشد.

تعریف (۱-۱-۱۰): یک راه[1] به طول k یک دنباله $v_k, e_k, \ldots, e_1, v_1, e_1, v_0$ از رأس‌ها و یالهاست به طوری که برای هر i، $(i=1,2,\ldots,k)$، $e_i = v_{i-1}v_i$.

1. *Walk*

تعریف (۱-۱-۱۱): یک گذر (ترایل) راهی است که هیچ یال تکراری نداشته باشد.

تعریف(۱-۱-۱۲): یک مسیر گذری است که هیچ رأس تکراری نداشته باشد.

ـ درجه گره W در گراف G را با $\deg G(W)$ نمایش داده و برابر است با تعداد یالهایی که در G با W، هموقوعند. در محاسبه درجه گره‌ها هر حلقه دوبار به حساب می‌آید. هر گره با درجه صفر را یک گره منزوی (ایزوله یا منفرد)، و هر گره با درجه ۱ را یک گره انتهایی گوییم.

تعریف (۱-۱-۱۳): گرافی که بین هر دو گره آن مسیری وجود داشته باشد، را گراف همبند می‌نامیم.

تعریف (۱-۱-۱۴): درخت گرافی است ساده، همبند و بدون دور.

ـ اگر T یک درخت و مجموعه رأسها و یالهای آن را به ترتیب با $E(T),V(T)$ نشان می‌دهیم، خواهیم داشت $E(T)=V(T)-1$.

تعریف (۱-۱-۱۵): هر زیر مجموعه از یالهای غیر همجوار را یک تطابق نامند.

برای آشنایی بیشتر با تعاریف قبل ومثال های آن می‌توان به مرجع [۱] رجوع کرد.

۱-۲ متغیرهای مختلط و کاربردها

معمولاً عدد مختلط (x,y) را به صورت زیر نمایش می‌دهند.

$$z=(x,y) \qquad (۱-۱)$$

به علاوه اعداد حقیقی y,x را، به ترتیب، قسمت‌های حقیقی و موهومی z می‌نامند و می‌نویسند

$$\operatorname{Re} Z = x, \quad \operatorname{Im} Z = y \tag{۲-۱}$$

و اگر عدد حقیقی x را به صورت $(x,0)$ یا $(x,0)$ بگیریم و i نمایش عدد موهومی محض $(0,1)$ باشد، واضح است که

$$Z = x + iy \tag{۳-۱}$$

و ملاحظه می‌کنیم

$$i^2 = (0,1)(0,1) = (-1,0) = -1$$

تعریف(۱-۲-۱): تابع f از متغیر مختلط z در یک مجموعه باز تحلیلی است اگر در هر نقطه آن مجموعه دارای مشتق باشد. اگر از تابع fای سخن به میان می‌آوریم که در مجموعه S تحلیلی است و S باز نیست منظور این است که f در مجموعهٔ بازی شامل S تحلیلی است. به خصوص f در نقطه z_0 تحلیلی است اگر در یک همسایگی z_0 تحلیلی باشد.

مثال (۱-۲-۲): تابع $f(z) = \frac{1}{z}$ در هر نقطه ناصفر صفحه تحلیلی است. اما تابع $f(z) = |z|^2$ در صفر تحلیلی نیست زیرا مشتق آن در $z = 0$ موجود است ولی در هیچ همسایگی از صفر تحلیلی نیست.

قضیه (۱-۲-۳):(قضیه کوشی گورسا)

اگر تابع f در همه نقاط درون و روی مسیر ساده بسته C تحلیلی باشد، آنگاه

$$\int_C f(z)dz = 0 \tag{۴-۱}$$

قضیه (۱-۲-۴): فرض کنید f همه جا درون و بر مسیر ساده بسته C، که در جهت مثبت

گرفته شده است، تحلیلی باشد. اگر z_0 نقطه دلخواهی درون C باشد، آنگاه

$$f(z_0)=\frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}\,dz \qquad (۵-۱)$$

قضیه (۱-۲-۵): اگر تابع f در سراسر حوزه D پیوسته باشد و به ازای هر مسیر بسته C واقع در D،

$$\int_C f(z)dz=0 \qquad (۶-۱)$$

آنگاه f در سراسر D تحلیلی است.

۱-۳ سری تیلر

قضیه (۱-۳-۱): فرض کنید f در سراسر قرص باز $|z-z_0|<R$ به مرکز z_0 و شعاع R_0 تحلیلی باشد. در این صورت در هر نقطه z در آن قرص، $f(z)$ دارای نمایش سری

$$f(z)=\sum_{n=\cdot}^{\infty}a_n(z-z_{\cdot})^n \quad \left(|z-z_{\cdot}|<R_{\cdot}\right) \qquad (|Z-Z_0|<R_0) \qquad (۷-۱)$$

است که در آن

$$a_n=\frac{f^{(n)}(z_0)}{n!} \qquad (۸-۱)$$

یعنی هرگاه $|z-z_0|<R$، این سری توانی به $f(z)$ همگرا است.

۴-۱ سری لوران

اگر تابع f در نقطه z_0 تحلیلی نباشد، نمی‌توانیم از قضیه تیلر در آن نقطه استفاده کنیم. ولی اغلب می‌توان یک نمایش سری برای $f(z)$ شامل توان‌های مثبت و منفی $z-z_0$ پیدا کرد.

قضیه (۱-۴-۱): فرض کنید تابع f در سراسر حوزه طوقی $R_1<|z-z_0|<R_2$ تحلیلی و C معرف مسیر ساده بسته‌ای در جهت مثبت حول z_0 و واقع در این حوزه باشد. در این صورت در هر نقطه z از آن حوزه، $f(z)$ دارای نمایش سری

$$f(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n+\sum_{n=1}^{\infty}\frac{b_n}{(z-z_0)^n}\qquad \left(R_1<|z-z_0|<R_2\right)\qquad (۱-۹)$$

است که در آن

$$a_n=\frac{1}{2\pi i}\int_C\frac{f(z)dz}{(z-z_0)^{n+1}}\qquad (n=0,1,2,...)\qquad (۱-۱۰)$$

$$b_n=\frac{1}{2\pi i}\int_C\frac{f(z)dz}{(z-z_0)^{-n+1}}\qquad (۱-۱۱)$$

(a_n و b_n ضرایب مختلط ثابتی می باشند.)

رابطه (۱-۹) را اغلب به صورت زیر می‌نویسند:

$$f(z)=\sum_{n=-\infty}^{\infty}C_n(z-z_0)^n\qquad \left(R_1<|z-z_0|<R_2\right)\qquad (۱-۱۲)$$

که در آن

$$C_n=\frac{1}{2\pi i}\int_C\frac{f(z)dz}{(z-z_0)^{n+1}}\qquad (n=0,\pm1,\pm2,...)\qquad (۱-۱۳)$$

هر یک از سری‌های (۱-۹) و (۱-۱۲) را سری لوران می‌نامند.