مجلهٔ پژوهش فیزیک ایران، جلد ۲۱، شمارهٔ ۲، تابستان ۱۴۰۰
DOI: 10.47176/ijpr.21.2.51066

خوشه‌یابی پیمانگی در سری‌های زمانی مالی

دانیال پاپی و سید محمدصادق موحد

دانشکدهٔ فیزیک، دانشگاه شهید بهشتی، تهران

پست الکترونیکی: m.s.movahed@ipm.ir

(دریافت مقاله: ۱۳۹۹/۰۳/۱۱ ؛ دریافت نسخهٔ نهایی: ۱۳۹۹/۱۰/۱۸)

چکیده

در این مقاله با تکیه بر خوشه‌یابی در شبکه‌های پیچیده که می‌تواند ویژگی‌های بزرگ مقیاس شبکه را تعیین کند، به مطالعهٔ ۴۸ بازار مالی در سراسر دنیا می‌پردازیم. برای این منظور روش بیشینه‌سازی پیمانگی را برای شبکه‌های جهت‌دار و وزن‌دار توسعه می‌دهیم. با کمک معیار همبستگی خطی، ماتریس مجاورت را تشکیل داده و با استفاده از نظریهٔ ماتریس‌های تصادفی فضای ویژه مقداری ماتریس خود را به دو بخش نامربوط و مربوط تقسیم‌بندی می‌کنیم. با در نظرگرفتن پنجرهٔ زمانی و تحول آن در طول سری‌های زمانی، نتایج ما نشان می‌دهد که در حوالی بحران‌های مالی، خوشه‌هایی که غالباً تحت تاثیر ویژگی‌های جغرافیایی است، تشکیل می‌شوند و از منظر شبکه‌های پیچیده، کاتوره‌ای‌ترین رفتار خود را نشان می‌دهند.

واژه‌های کلیدی: فیزیک اقتصاد، شبکهٔ پیچیده، خوشه‌یابی، بیشینه‌سازی پیمانگی، نظریهٔ ماتریس‌های تصادفی

۱. مقدمه

گراف‌ها، تصویر انتزاعی مفیدی را از ساختار بسیاری از شبکه‌ها، فراهم می‌کنند [۱]. این شبکه‌ها می‌توانند از شبکه‌های اجتماعی و شبکه‌های کامپیوتری تا شبکه‌های زیستی یا حتی اقتصادی تشکیل شده باشند. تلاش‌های بسیاری برای تشخیص ساختار هندسی شبکه‌های مختلف شده است [۲ و ۳]. بازهٔ گسترده‌ای از ابزار و رهیافت‌ها وجود دارند که می‌توانند به ما در فهم ساختار کوچک مقیاس شبکه‌ها کمک کنند. برای مثال می‌توان به همبستگی، ارتباط‌ها[۱] یا الگوی بازگشتی[۲] اشاره کرد. اما برای فهم ساختار مقیاس متوسط یا مقیاس بزرگ به منظور درک یک تصویر باز از شبکه‌ها، موارد مطرح شده کافی نیستند [۴].

در دههٔ گذشته علاقهٔ زیادی هم در مطالعهٔ شهودی شبکه‌ها و هم در توسعهٔ ابزارهای ریاضی و محاسباتی برای به دست آوردن بینشی بهتر از داده‌های شبکه به وجود آمده است. یکی از رهیافت‌های اساسی برای مطالعهٔ شبکهٔ تمرکز بر ویژگی‌های تک تک گره‌ها یا گروه کوچکی از آنهاست. به همین منظور سؤالاتی نظیر: مهم‌ترین گره در شبکه کدام است؟ یا قوی‌ترین

۱. Connections

۲. Recurrent patterns

ارتباط در شبکه کدام است؟ از نمونه سؤالاتی هستند که پاسخ به آنها می‌تواند اطلاعات کمّی در مورد ساختار بزرگ مقیاس شبکه‌های مورد مطالعه در اختیار ما قراردهد. یکی از کامل‌ترین مطالعات انجام شده برای ساختار بزرگ مقیاس در شبکه‌ها، خوشه‌یابی[۱] یا تشخیص جوامع[۲] است [۵]. در این مطالعه، یک جامعه یک زیرشبکهٔ متراکم داخل شبکهٔ بزرگ‌تر است، مثل گروه صمیمی‌ای از دوستان در یک شبکهٔ اجتماعی. اگرچه که این جوامع تنها شکل مورد علاقه از ساختار بزرگ مقیاس نیستند اما آنها می‌توانند ویژگی کلی تحقیقات را به ما نشان دهند. از این جهت اولویت اول در بررسی ما هستند.

مسئلهٔ تشخیص جامعه و گروه در شبکه‌ها به دلیل خوش تعریف نبودن، مسئلهٔ چالش برانگیزی است. اما باید خاطرنشان کرد که بر سر این موضوع که مسئله در مورد پیدا کردن مناطق متراکم‌تر در شبکه است، توافقی نسبی وجود دارد اما متاسفانه این کار یک روش مشخصی ندارد. گسترهٔ زیادی از تعاریف برای جامعه بر اساس ویژگی خاص شبکه تعریف شده‌اند و با انتخاب هرکدام از این تعاریف، مجموعه‌ای از الگوریتم‌های متناسب با آن تعریف نیز وجود دارند [۵]. همواره این سؤال مطرح است که: کدام یک از این الگوریتم‌ها برای پیدا کردن جوامع متناسب با تعریف خود، عملکرد مناسبی دارند؟ برای خارج شدن از این چرخه، معمولاً به یکی از دو راه حل زیر مراجعه می‌شود:

راه حل اول: شبکه‌ای واقعی را که با استفاده از ابزارهای غیر شبکه‌ای نسبت به جوامع آن دید داریم، انتخاب می‌کنیم و سپس الگوریتم خود را روی شبکه می‌آزماییم. اگر نتایج به دست آمده با انتظار اولیهٔ ما سازگار باشد، آن الگوریتم موفق بوده است [۶].

راه حل دوم: یک شبکهٔ ساختگی با ساختاری از جوامع داخل آن را توسط رایانه شبیه‌سازی می‌کنیم و الگوریتم را روی آن بررسی می‌کنیم [۷–۱۰]، اگرچه این رهیافت‌ها روش ملموسی را برای بررسی عملکرد الگوریتم‌های تشخیص جامعه پیشنهاد می‌دهند اما همواره جای این سؤال باقی می‌ماند که آیا این عملکرد خوب در مقابل داده دنیای واقعی نیز خوب باقی می‌ماند؟

با عنایت به پیچیدگی‌های موجود در خصوص تعیین جامعه و خوشگی در شبکه‌های پیچیده، روش‌هایی از جمله بخش‌بندی طیفی[۱۱] و الگوریتم کرنیگهان- لین[۳] [۱۲] معرفی شده است. از منظر فیزیک سیستم‌های پیچیده، روش‌های بهینه‌سازی با تکیه بر کمینه کردن انرژی و یافتن حالت پایهٔ سیستم، نظیر روش E/I [۱۳] ، روش مبتنی بر راست‌نمایی[۴] [۱۴] و سایر روش‌ها [۱۵] رایج هستند، اما گسترده‌ترین روشی که می‌توان از آن یاد کرد، به اسم پیمانگی[۵][۱۶ و ۱۷] شناخته می‌شود. البته توجه به روش‌های بنیادی‌تر سبب شد مدل بلوک نیز ارائه شود [۱۸ و ۱۹]. علی‌رغم مطالعات متعددی که در خصوص مفهوم خوشگی انجام شده است، رویکرد واحدی برای تعیین آن وجود ندارد. برای مثال در تشکیل ساختارهای بزرگ مقیاس کیهانی، خوشگی مطابق با رویکرد تشکیل ساختار سلسله‌مراتبی مورد توجه است [۲۰]. در این گونه تحلیل‌ها با عنایت به معیارهای مختلف همبستگی مانند همبستگی دونقطه‌ای بدون وزن، خوشگی مطالعه شده است [۲۱]. برای مطالعهٔ خواص ترمودینامیکی سیستم‌های واقعی با احتساب اندرکنش بین اجزای تشکیل دهندهٔ آنها، نیز روش‌هایی از جمله بسط خوشه‌های مایر معرفی شده است. این روش در مطالعهٔ سیستم‌ها با تعداد زیادی اجزای تشکیل دهنده سابقهٔ طولانی دارد [۲۲]. به طور مشخص برای یافتن خوشه‌ها و جامعه‌ها در شبکه‌های داده‌های مالی می‌توان به روش‌های زیر اشاره کرد: روش خوشه‌یابی فازیبرای پیش‌بینی بازارهای مالی که در مرجع [۲] آمده است، روش K_means که به طور مشخص برای هوشمندسازی بازاریابی یا حتی معرفی محصولات متناسب با مشتری در بازاریابی برخط استفاده شده است [۲۳–۲۵]، روش سلسه‌مراتبی که معمولا برای پیش‌بینی کوتاه‌مدت قیمت بعد از انتشار گزارش در یک بازار مالی به کار می‌رود [۲۶]. برای تکمیل موارد به استفاده از شبکهٔ عصبی مصنوعی برای یافتن

۱. Clustering

۲. Community detection

۳. Kernighan–Lin

۴. likelihood-based

۵. Modularity

خوشه‌ها، نیز اشاره می‌کنیم [۲۷]. اهمیت یافتن خوشه‌ها در شبکه‌های مالی، به طور کلی به مواردی همچون مدیریت خطر در بازارهای مالی، پیش‌بینی بحران‌های اقتصادی، تعیین گذار بین حالت‌های مختلف بازارها نیز مربوط می‌شود [۲۸]. در این مقاله به منظور بررسی خوشگی در شبکهٔ بازارهای مالی ۴۸ کشور مختلف و در نظر گرفتن آن که همبستگی‌های وزن‌دار و حتی منفی بین اقتصاد دو کشور نیز شامل اطلاعات مفیدی از خوشگی است، از الگوریتم وزن‌دار شدهٔ بیشینه‌سازی پیمانگی استفاده خواهیم کرد و معیار پیمانگی را به عنوان نتیجه‌ای از خوشگی قوی یا ضعیف در حوالی بحران‌های مالی بررسی خواهیم کرد. به همین منظور ابتدا در بخش ۲ به توصیف بیشینه‌سازی پیمانگی و روش حل آن بر روی شبکه‌ها می‌پردازیم سپس در بخش ۳، این روش را برای شبکه‌های وزن‌دار توسعه می‌دهیم. در بخش ۴ بعد از فیلتر داده‌ها توسط نظریهٔ ماتریس‌های تصادفی، به خوشه‌بندی این شبکه بر اساس الگوریتم وزن‌دار خواهیم پرداخت. پیشنهادها برای ادامه مقاله را در بخش آخر ارائه خواهیم کرد.

۲. بیشینه‌سازی پیمانگی برای شبکه‌های بدون وزن

در روش‌های تشخیص اجتماع، عموماً شبکه را تشکیل شده از چند زیرگروه ذاتی در نظر می‌گیریم و وظیفهٔ تحلیل‌گر پیدا کردن این زیرگروه‌ها است. تعداد و اندازهٔ این گروه‌ها معمولاً توسط خود شبکه تعیین می‌شود. در این بخش روش پیمانگی برای یافتن خوشه را معرفی می‌کنیم [۵]. پیمانگی در حقیقت معیاری از تشابه اعضا برای قرار گرفتن در یک گروه است. این معیار با مقایسهٔ بین جمع تعداد یال‌های داخلی یک گروه با جمع این یال‌ها در حالتی که شبکه به صورت کاملاً اتفاقی باشد، به دست می‌آید. پیمانگی می‌تواند مقداری مثبت یا منفی داشته باشد که مقدار مثبت آن بیانگر احتمال وجود یک ساختار جامعه‌ای است. بنابراین می‌توان با استفاده از ایدهٔ کمینه کردن انرژی در فیزیک، کاری برعکس انجام داد یعنی برای پیدا کردن جوامع همه حالت‌های ممکن را بررسی کرد و سپس جوابی که بیشترین مقدار مثبت پیمانگی را ارائه می‌دهد به عنوان ساختار خوشگی در شبکه در نظر بگیریم. فرض کنید شبکهٔ ما از n گره تشکیل شده است. برای مسئلهٔ خاص تقسیم شبکه به دو گروه، $s_i = 1$ اگر گره i متعلق به گروه ۱ باشد و $s_i = -1$ اگر این گره متعلق به گروه ۲ باشد. تعداد یال‌های بین دو رأس را توسط اعضای ماتریس مجاورت A_{ij} بیان می‌کنیم که معمولاً ۰ یا ۱ هستند (در مورد شبکه‌های وزن‌دار در ادامه صحبت می‌شود). مقدار انتظاری یال‌های بین دو گره i و j در حالت کاملاً کاتوره‌ای برابر $k_i k_j / 2m$ است که k_i و k_j درجات گره‌های مورد نظر و $m = \frac{1}{2}\sum k_i$ برابر مجموع یال‌ها در شبکه است. بنابراین معیار پیمانگی Q از جمع مقدار $A_{ij} - k_i kj / 2m$ روی تمامی جفت گره‌های i و j که در یک گروه قرار دارند، به دست می‌آید. باید دقت کرد که مقدار $\frac{1}{2}(s_i s_j + 1)$ برای زمانی که دو گره در یک گروه باشند برابر یک و برای زمانی که این دو گره از دو گروه متفاوت باشند برابر صفر خواهد بود. همچنین باید اشاره کرد که جمع درایه‌های هر سطر یا ستون ماتریس $A_{ij} - k_i kj / 2m$ برابر صفر خواهد بود. در نتیجه می‌توان شکل کلی پیمانگی را اینگونه نمایش داد [۲۹]:

$$Q = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i kj}{2m}\right)\left(s_i s_j + 1\right) = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i kj}{2m}\right)s_i s_j, \quad (1)$$

رابطهٔ ۱ را می‌توان به شکل ماتریسی زیر بازنویسی کرد:

$$Q = \frac{1}{4m} s^T B s, \quad (2)$$

که در آن s یک بردار ستونی با اعضای s_i است و اعضای ماتریس حقیقی و متقارن B به شکل زیر نوشته شده‌اند:

$$B_{ij} = A_{ij} - \frac{k_i kj}{2m}, \quad (3)$$

که آن را ماتریس پیمانگی می‌نامیم. با توجه به صفر بودن جمع درایه‌های سطر یا ستون ماتریس B، این ماتریس ویژه‌برداری به شکل $(1,1,1,1,\ldots)$ با ویژه‌مقدار صفر خواهد داشت. این مقدار یادآور ماتریسی است که به اسم گراف لاپلاسی شناخته

می‌شود[۳۰] و پایه و اساس یکی از بهترین روش‌های شناخته شده بخش‌بندی گراف، یعنی بخش‌بندی طیفی است [۱۱ و ۳۱] و همان ویژگی‌ها را دارد. با استفاده از رابطهٔ ۲ و بسط s به شکل ترکیب خطی‌ای از ویژه مقادیر بهنجار شده B یعنی $s=\sum_{i=1}^{n} a_i u_i$ و $a_i = u_i^T.s$ می‌توانیم رابطهٔ زیر را بنویسیم:

$$Q=\frac{1}{4m}\sum_{i}\left(a_i u_i\right)^T B\sum_{j} a_j u_j = \frac{1}{4m}\sum_{i=1}^{n}\left(u_i^T.s\right)^2 \beta_i, \quad (۴)$$

که در آن β_i ویژه مقادیر B متناظر با u_i است. ویژه مقادیر را به شکل کاهشی شماره‌گذاری می‌کنیم، یعنی $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_n$ قصد ما بیشینه کردن مقدار پیمانگی با انتخاب مناسب بخش‌بندی شبکه است. برای این کار باید موازی‌ترین حالت ممکن را که برابر با بیشینه کردن $u_1^T.s$ است، انتخاب کنیم. این کار ساده است. برای بیشینه شدن این مقدار کافی است، در حالتی که مولفهٔ متناظر s_i از ویژه‌بردار u_1 مثبت بود، s_i را برابر یک و در حالتی که مؤلفهٔ متناظر منفی بود، s_i را برابر منفی یک در نظر بگیریم. به بیان دیگر تمام گره‌هایی که متناظر با مولفه‌های مثبت هستند را در یک گروه و بقیه را در گروه دیگر قرار می‌دهیم. الگوریتم ذکر شده نه تنها قابلیت تقسیم یک شبکه به شکل موثر را دارد بلکه زمانی که تقسیم‌بندی خوبی وجود ندارد با به دست آوردن مقدار غیر مثبت پیمانگی از بخش‌بندی شبکه خودداری می‌کند. برای تقسیم شبکه به بیش از دو بخش باید اشاره شود که ما مجاز به پاک کردن یال‌های بین هردو گروه پس از تقسیم بندی آنها و انجام دوبارهٔ این الگوریتم روی زیرگراف‌های به دست آمده، نیستیم. رهیافت درست تعریف کمیت ΔQ برای زیرگروه به دست آمدهٔ g با اندازهٔ n_g به شکل زیر است:

$$\Delta Q=\frac{1}{2m}\left[\frac{1}{2}\sum_{i.j\in g} B_{ij}\left(s_i s_j+1\right)-\sum_{i.j\in g} B_{ij}\right], \quad (۵)$$

$$\Delta Q=\frac{1}{4m}\left[\sum_{i.j\in g} B_{ij} s_i s_j-\sum_{i.j\in g} B_{ij}\right],$$

$$\Delta Q=\frac{1}{4m}\sum_{i.j\in g}\left[B_{ij}-\delta_{ij}\sum_{k\in g} B_{ik}\right] s_i s_j,$$

$$\Delta Q=\frac{1}{4m} s^T B^g s,$$

که در آن δ_{ij} همان دلتای کرونکر و $B^{(g)}$ یک ماتریس $n_g \times n_g$ عضو است که به شکل زیر تعریف شده است:

$$B_{ij}^{(g)}=B_{ij}-\delta_{ij}\sum_{k\in g} B_{ik}, \quad (۶)$$

به خاطر یکی بودن شکل روابط ۲ و ۵، می‌توانیم برای بیشینه کردن ΔQ همان رهیافت طیفی را این بار برای ماتریس پیمانگی $B^{(g)}$ استفاده کنیم. باید توجه کرد که جمع سطر و ستون ماتریس $B^{(g)}$ همچنان صفر است و ΔQ نیز برای حالتی که گروه g تقسیم ناپذیر است برابر صفر می‌شود.

حال می‌توان برای پیدا کردن بهترین حالت تقسیم بندی از ترکیب روش دیگری با روش بیشینه‌سازی پیمانگی استفاده کرد. فرض کنید ما تقسیم‌بندی اولیه‌ای برای بخش‌بندی گره‌های شبکه به دو گروه، در اختیار داریم و سپس به دنبال گره‌ای می‌گردیم که اگر بین این دو گروه جابه‌جا شد، بزرگ‌ترین تغییر مثبت در پیمانگی را باعث شود یا اگر تغییر مثبتی وجود نداشت کمترین تغییر منفی را ایجاد کند. این جابه‌جایی را مرتباً همراه با این قید که هر گره تنها یک بار اجازه جابه‌جا شدن دارد، تکرار می‌کنیم. زمانی که همهٔ n گره جابه‌جا شدند، از بین حالات میانی شبکه در حین انجام این الگوریتم به دنبال حالتی می‌گردیم که بیشترین پیمانگی را داشته باشد و آن حالت را انتخاب می‌کنیم. دوباره از حالت انتخاب شده شروع کرده و کل فرآیند را تا زمانی تکرار می‌کنیم که دیگر هیچ پیشرفتی برای پیمانگی به وجود نیاید. این روش بسیار شبیه به الگوریتم کرنیگهان-لین [۱۲] است.

علی‌رغم سادگی، ترکیب بیشینه‌سازی پیمانگی شرح داده شده با این روش و جابه‌جا کردن گره‌ها در جوامع برای به دست آوردن بهترین مقدار پیمانگی، می‌تواند منجر به نتایج بسیار مناسبی شود.

در بخش بعدی این الگوریتم بیشینه‌سازی را برای شبکه‌های وزن‌دار و علامت‌دار بازنویسی خواهیم کرد.

۳ بیشینه‌سازی پیمانگی برای شبکه‌های وزن‌دار.

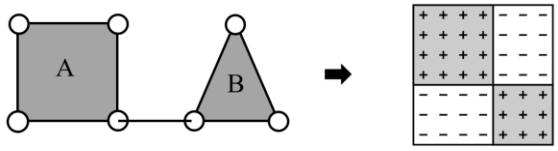
در این بخش الگوریتم شرح داده شده در بخش ۲ را متناسب با یک شبکهٔ وزن‌دار بازنویسی خواهیم کرد. در ابتدا ماتریس مجاورت شبکهٔ بی‌وزن بخش قبلی را با ماتریس مجاورت شبکهٔ وزن‌دار که در آن هر عضو این ماتریس می‌تواند عددی بین ۱ تا ۱- باشد، جایگزین می‌کنیم. با توجه به موارد گفته شده، می‌توان رابطهٔ ۱ را به شکل زیر بازنویسی کرد:

$$Q=\frac{1}{2\omega}\sum_{ij}\left(\omega_{ij}-\frac{\omega_i\omega_j}{2\omega}\right)\delta\left(C_i,C_j\right), \qquad (۷)$$

در این رابطه ω_{ij} وزن یال اتصالی بین دو رأس $i\,j$ ، ω_i ω_j درجات این دو رأس و $\omega=\frac{1}{2}\sum\omega_i$ است. در این رابطه C_i بیانگر گروهی است که گره i به آن تعلق دارد. پیمانگی Q بر اساس ماتریس مجاورت وزن‌دار ω_{ij} بیان شده است که عدد مربوط به هر عضو این ماتریس بیانگر قدرت اتصال بین دو گره i و j است.

برای اثبات عملکرد نادرست رابطهٔ ۷ در تشخیص ساختار جوامع شبکه، مثالی در نظر می‌گیریم. فرض کنید با یک شبکه با ساختار ایدهال جوامع مانند شکل ۱ سر و کار داریم. برای ساده‌سازی ریاضی گره‌های موجود در جامعه A را در حالت ۱+ و گره‌های موجود در جامعه B را در حالت ۱- فرض می‌کنیم.

سؤال اصلی‌ای که مطرح است این است که آیا جوامع A و B را می‌توان از ماتریس مجاورت داده‌های بیان شده، به دست آورد؟ جواب صریح این است که با پیمانگی معرفی شده در رابطهٔ ۷، این امکان وجود ندارد. برای پیدا کردن ساختار جوامع از ماتریس مجاورت در شبکهٔ شکل ۱، لازم است که فرمول‌بندی معادلهٔ ۷ را بازنویسی کنیم. فرض کنید یک شبکهٔ پیچیده بدون جهت و وزن‌دار با وزن ω_{ij} مثل بالا در اختیار داریم. قدرت اتصال یک گره به صورت زیر تعریف می‌شود:



شکل ۱. مثالی از بررسی شبکهٔ وزن‌دار [۳۲].

$$p_i=\frac{\omega_i}{2\omega}, \qquad (۸)$$

می‌توانیم این مقدار را به عنوان احتمال اتصال یک گره به گره‌های دیگر در حالتی که این شبکه کاتوره‌ای چیده شده باشد، در نظر گرفت. این دقیقاً همان عبارتی است که در بخش قبل نیز برای محاسبهٔ پیمانگی به شکل زیر به کار برده شده بود:

$$p_i p_j=\frac{\omega_i\omega_j}{(2\omega)^2}, \qquad (۹)$$

وزن منفی که در ابتدا معرفی کردیم می‌تواند مفهوم احتمالی p_i را از بین ببرد به طوری که دیگر الزاماً مقدار p_i بین صفر و یک نخواهد بود. مشکل این فرضیه از آنجا نشأت می‌گیرد که برای یال‌های با وزن مثبت و منفی بین گره‌ها تنها یک احتمال یکتا در نظر می‌گیریم. به هر حال اگر فرض کنیم دو احتمال متفاوت برای تشکیل یال‌ها وجود دارد، یکی برای وزن مثبت و دیگری برای وزن منفی، مشکل از بین می‌رود. برای فرمول‌بندی این رهیافت، ابتدا قصد داریم وزن‌های مثبت و منفی را از هم جدا کنیم، یعنی :

$$\omega_{ij}=\omega_{ij}^{+}-\omega_{ij}^{-}, \qquad (۱۰)$$

در رابطهٔ ۱۰ عبارت ω_{ij}^{+} به معنی مقادیر مثبت و عبارت ω_{ij}^{-} به معنی مقادیر منفی از ماتریس مجاورت است، همچنین وزن مثبت و منفی توسط روابط زیر به دست می‌آیند:

$$\omega_i^{+}=\sum_j\omega_{ij}^{+}\ .\ \omega_{ij}^{-}=\sum_j\omega_{ij}^{-}, \qquad (۱۱)$$

و مقدار کل وزن مثبت و منفی برابر است با:

$$2\omega^{+}=\sum_i\omega_i^{+}=\sum_{ij}\omega_{ij}^{+}\quad .\quad 2\omega^{-}=\sum_i\omega_i^{-}=\sum_{ij}\omega_{ij}^{-}, \qquad (۱۲)$$

با این تعاریف، احتمالات مرتبط با وزن‌های مثبت و منفی مانند زیر خواهند بود:

جدول ۱. جدول مربوط به اسامی کشورها و اعداد مختص به هر شاخص در داده.

کشور	شاخص	کشور	شاخص	کشور	شاخص
PAK	۳۳	HK	۱۷	ARG	۱
PER	۳۴	HUN	۱۸	AUS	۲
POL	۳۵	IND	۱۹	AUT	۳
PRT	۳۶	IDN	۲۰	BEL	۴
RUS	۳۷	IRL	۲۱	BRA	۵
SGP	۳۸	ISR	۲۲	CAN	۶
ZAF	۳۹	ITA	۲۳	CHL	۷
ESP	۴۰	JPN	۲۴	CHN	۸
LKA	۴۱	JOR	۲۵	COL	۹
SWE	۴۲	KOR	۲۶	CZE	۱۰
CHE	۴۳	MYS	۲۷	DNK	۱۱
TW	۴۴	MEX	۲۸	EGY	۱۲
THA	۴۵	MAR	۲۹	FIN	۱۳
TUR	۴۶	NLD	۳۰	FRA	۱۴
UK	۴۷	NZL	۳۱	GRC	۱۵
USA	۴۸	NOR	۳۲	DEU	۱۶

$$p_i^+ = \frac{\omega_i^+}{۲\omega^+} \quad . \quad p_i^- = \frac{\omega_i^-}{۲\omega^-}, \tag{۱۳}$$

اکنون دو عبارت برای تشکیل پیمانگی در اختیار داریم: اولی اختلاف وزن واقعی را با مدل تهی شبکهٔ اتفاقی که توسط احتمال p_i^+ تشکیل می‌شود، به ما می‌دهد و دومی نیز این کار را برای وزن‌های منفی محاسبه می‌کند. پس تعاریف زیر می‌توانند سودمند باشند:

$$\begin{aligned} Q^+ &= \frac{۱}{۲\omega^+}\sum_{ij}\left(\omega_{ij}^+ - \frac{\omega_i^+\omega_j^+}{۲\omega^+}\right)\delta\left(C_iC_j\right), \\ Q^- &= \frac{۱}{۲\omega^-}\sum_{ij}\left(\omega_{ij}^- - \frac{\omega_i^-\omega_j^-}{۲\omega^-}\right)\delta\left(C_iC_j\right), \end{aligned} \tag{۱۴}$$

پیمانگی کل در نهایت از تقابل دو پیمانگی بالا پدید می‌آید. باید توجه کرد زمانی تعداد مشخصی از اعضا شرایط مناسبی برای تشکیل یک گروه دارند که مقدار Q^+ این اعضا بیشتر از مقدار Q^- آنها باشد. پس اگر ما تمایل داشته باشیم تا پیمانگی کل را با استفاده از این دو پیمانگی مثبت و منفی بسازیم، این مقدار را می‌توانیم به شکل زیر بیان و ساده کنیم [۳۲]:

$$\begin{aligned} Q &= \frac{۲\omega^+}{۲\omega^+ + ۲\omega^-}Q^+ - \frac{۲\omega^-}{۲\omega^+ + ۲\omega^-}Q^-, \\ Q &= \frac{۱}{۲\omega^+ + ۲\omega^-}\sum_{ij}\left[\omega_{ij} - \left(\frac{\omega_i^+\omega_j^+}{۲\omega^+} - \frac{\omega_i^-\omega_j^-}{۲\omega^-}\right)\right]\times \\ &\delta\left(C_iC_j\right), \end{aligned} \tag{۱۵}$$

در بخش بعدی به مطالعهٔ ۴۸ بازار مالی در نظر گرفته شده با استفاده از روش مرسوم بیشینه‌سازی پیمانگی و روش وزن‌دار شدهٔ آن خواهیم پرداخت.

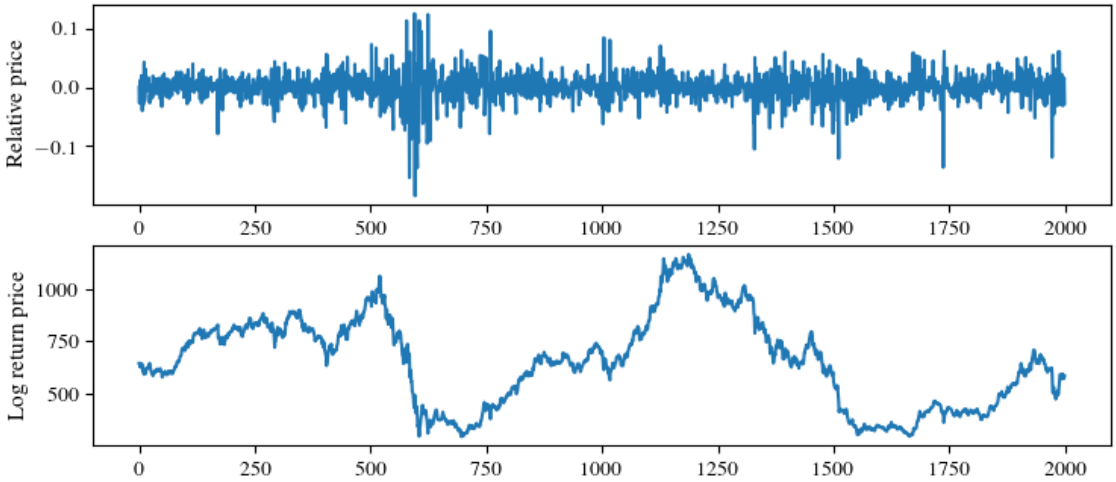
۴. اعمال روش خوشه‌یابی پیمانگی بر روی داده‌های بازارهای سهام

در این فصل قصد داریم با کمک الگوریتم ارائه شده در بخش‌های قبل و به کارگیری روش‌های آماری که در ادامه ذکر خواهیم کرد، به کشف ساختار بزرگ مقیاس (ساختار جوامع) در شبکهٔ اقتصادی منتخبی از کشورهای دنیا بپردازیم و درستی روش پیشنهادی را راستی‌آزمایی کنیم. برای این کار ابتدا داده‌های مورد استفاده را شرح خواهیم داد و بعد از معرفی چند روش برای از بین بردن نوفهٔ داده‌ها، به چگونگی عملکرد رهیافت ارائه شده در بخش ۲ و ۳ بر روی این داده‌ها خواهیم پرداخت.

۱.۴. شرح داده‌های مورد مطالعه

دادهٔ مورد استفاده در این مدل‌سازی شامل ۴۸ بازار سهام توسعه یافته یا در حال توسعه از کشورهای مختلف است که توسط MSCI [۱] ساخته شده است. جدول ۱ برای نشان دادن این ۴۸

۱. Morgan Stanley Capital International



شکل ۲. به ترتیب سری زمانی اصلی و سری زمانی بعد از لگاریتم برگشتی برای داده‌های آرژانتین.

کشور و شمارهٔ شاخص مورد استفاده برای آنها قرار گرفته است. برای مدل‌سازی، داده‌های روزانه این بازارها را در طی دورهٔ زمانی ۱۹۹۵ تا ۲۰۱۹، که شامل ۶۵۱۲ داده به ازای هر شاخص بازار می‌شود، استفاده کرده‌ایم.

دسته‌بندی MSCI برای این شاخص‌ها بر اساس سه معیار: توسعه یافتگی اقتصادی، اندازه و قابلیت نقد شوندگی و دسترسی به بازار انجام گرفته است و بر این اساس بازارهای ما به ۳ دستهٔ توسعه یافته ، نوظهور[1] و بازارهای رو به رشد[2] دسته‌بندی می‌شود. در این پایگاه داده ۲۳ بازار توسعه یافته، ۲۱ بازار نوظهور و ۴ بازار رو به رشد وجود دارد. بازارهای توسعه یافته شامل کانادا، ایالات متحده آمریکا (از آمریکا)، اتریش، بلژیک، دانمارک، فنلاند، فرانسه، آلمان، ایرلند، اسرائیل، ایتالیا، هلند، نروژ، پرتغال، اسپانیا، سوئد، سوئیس، بریتانیا (از اروپا)، استرالیا، هنگ کنگ، ژاپن، نیوزیلند، سنگاپور (از قسمت اقیانوس آرام آسیا) هستند. بازارهای نوظهور شامل برزیل، شیلی، کلمبیا، مکزیک، پرو (از آمریکا)، جمهوری چک، مصر، یونان، مجارستان، لهستان، روسیه، آفریقای جنوبی، ترکیه (از اروپا، خاورمیانه و آفریقا)، چین، هند، اندونزی، کره، مالزی، تایوان و تایلند (از آسیا) می‌شود. بازارهای در حال رشد نیز شامل آرژانتین (از آمریکا)، مراکش (از آفریقا)، اردن (از خاورمیانه) و پاکستان (از آسیا) می‌شود. این داده‌ها قیمت نسبی شاخص‌های بازار سرمایه را نشان می‌دهد به طوری که عدد ۱۰۰ بیانگر اولین روز مشاهده است [۳۳].

۲.۴. آماده‌سازی داده‌ها

برای کمتر کردن اثرات تورمی، واحد پولی که با آن قیمت را بیان می‌کنیم و البته برای بهتر نمایان شدن نوسانات قیمتی با استفاده از تعریف لگاریتم برگشتی در رابطهٔ ۱۶، مقیاس قیمتی جدید را ایجاد می‌کنیم.

$$r(t,\delta t) \equiv ln\, s(t+\delta t) - ln\, s(t) = ln \frac{s(t+\delta t)}{s(t)}\,, \qquad (۱۶)$$

به طور مثال، سری زمانی این قیمت نسبی برای بازار آرژانتین و این سری زمانی بعد از اعمال مقیاس قیمتی مناسب در شکل ۲ رسم شده است. مقیاس زمانی انتخاب شده برای این داده‌ها زمان معاملات است و دلیل آن نیز تعطیلی بازارهای سهام در تاریخ مشخص متفاوت در هر کشور است. این دلیل باعث می‌شود که در نظر گرفتن مقیاس زمانی فیزیکی عملا غیر ممکن باشد. لازم به ذکر است داده‌های ما شامل قیمت پایانی هر روز بازارهای سهام است و عملا بررسی این قیمت‌ها با مقیاس کوچک‌تر باعث مقایسهٔ نامناسب داده‌ها می‌شود زیرا اختلاف ساعتی کشورهای مختلف باعث حجم معاملات

۱. Emerging markets

۲. Frontier markets

متفاوت در ساعات مختلف روز می‌شود و می‌توان نشان داد که داده‌های کوچک‌تر از ۱ روز، ممکن است به مقایسه‌ای نامناسب منجر شود.

۳.۴. رهیافت استاندارد

قبل از آن که به بیان نتایج الگوریتم معرفی شده در بخش ۳ بپردازیم از رهیافتی استاندارد برای تحلیل داده‌های معرفی شده، استفاده می‌کنیم. در این رهیافت بعد از معرفی حد آستانه‌ای مناسب برای فیلتر کردن داده‌ها و البته دور ریختن اطلاعات مربوط به همبستگی‌های منفی از الگوریتم بخش ۲ استفاده می‌کنیم تا ساختار مربوط به شبکهٔ اقتصادی کشورها را استخراج کنیم.

با استفاده از نتایج کلی آماری [۳۴]، می‌توان به سادگی نشان داد برای دو سری زمانی X_i و X_j به طول T که شامل دو متغیر تصادفی مستقل با توزیع نرمال هستند، کمیت زیر:

$$z_{ij} \equiv artanhC_{ij} = \frac{۱}{۲} ln \frac{۱+C_{ij}}{۱-C_{ij}}, \qquad (۱۷)$$

تابع توزیعی با میانگین صفر دارد (C_{ij} ضریب همبستگی نمونه است). خطای استاندارد z_{ij} نیز به صورت زیر است:

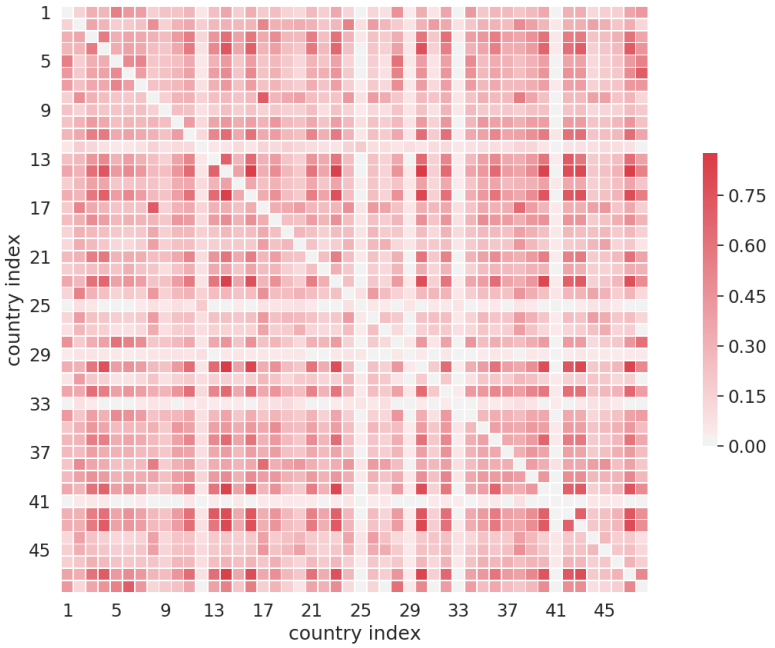
$$\sigma = (T-۳)^{-۱/۲}, \qquad (۱۸)$$

به بیانی دیگر با فرض بالا انتظار داریم مقادیر متغیرهای z_{ij} عموما با خطای استاندارد σ، حول و حوش صفر باشند. برای تشخیص انحرافات موثر از مدل بالا، می‌توان حد آستانهٔ τ را طوری تعریف کرد که فقط مقادیر خارج از خطای استاندارد τ، یعنی $|z_{ij}| > \tau\sigma$ در نظر گرفته شوند. به این ترتیب می‌توان آستانه را برای z_{ij} به صورت $z_\tau \equiv \tau\sigma$ در نظر گرفت. به بیانی دیگر همبستگی C_{ij} این مقدار حدی را می‌توان به شکل زیر نوشت:

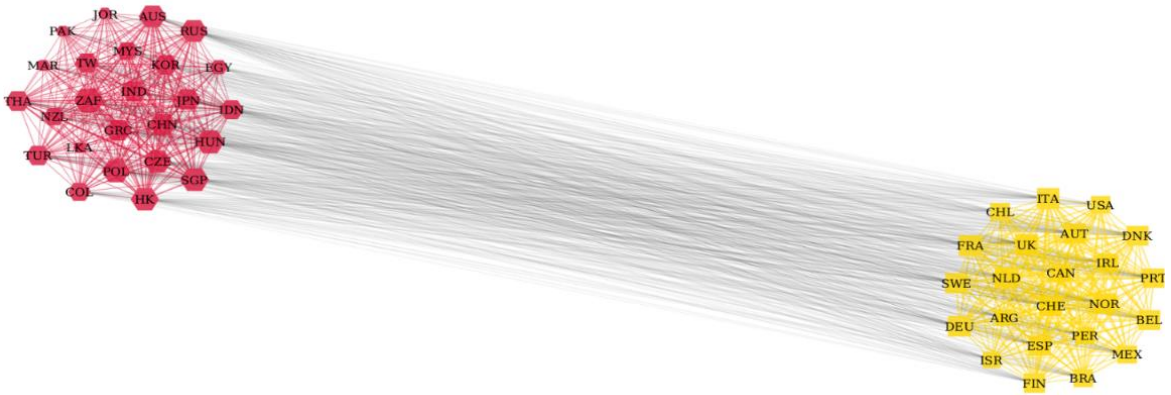
$$C_\tau \equiv tanhz_\tau = \frac{exp\left(\frac{۲\tau}{\sqrt{T-۳}}\right)-۱}{exp\left(\frac{۲\tau}{\sqrt{T-۳}}\right)+۱}, \qquad (۱۹)$$

اکنون می‌توانیم شبکهٔ مربوط به داده‌های اقتصادی و ساختار جوامع آن را استخراج کنیم. برای این کار ماتریس همبستگی متقابل داده‌ها را به دست می‌آوریم. با استفاده از حد آستانهٔ بخش قبل و انتخاب τ مناسب، ماتریس ضرایب همبستگی را فیلتر می‌کنیم و سپس با استفاده از تعریف اصلی بیشینه‌سازی پیمانگی بخش ۲، ساختار جوامع شبکه را استخراج می‌کنیم. برای فیلتر کردن ماتریس همبستگی، ابتدا باید به بخش قبل بازگردیم و τ مناسب را انتخاب کنیم. برای آن که عناصر ماتریس همبستگی را تا ۲ برابر انحراف از معیار فرضیهٔ معرفی شده در بخش قبل در نظر بگیریم مقدار τ را ۲ انتخاب کرده‌ایم. به این ترتیب با استفاده از عناصری از ماتریس همبستگی که $C_{ij} > C_\tau$ باشند شروع به ساخت شبکه خواهیم کرد. مقادیر ماتریس همبستگی بعد از اعمال فیلتر شرح داده شده و حذف عناصر قطری (به دلیل آن که عملاً تعریف ارتباط کامل هر کشور با خودش برای ما نامناسب است) برای کل بازهٔ داده‌های برگشتی لگاریتمی (از ۱۹۹۵ تا ۲۰۱۹) در شکل ۳ نمایش داده شده است.

حال می‌توان بدون نگرانی نسبت به یال‌های منفی از رابطهٔ اصلی بهینه سازی پیمانگی یعنی رابطهٔ ۱ برای استخراج جوامع مربوط به داده‌ها در این بازهٔ زمانی یاد شده استفاده کرد تا اطلاعاتی کلی از این روند را به دست آورده باشیم. شکل ۴ شبکهٔ استخراج شده از این الگوریتم را نشان می‌دهد. در این شکل یال‌های رنگی، یال‌های داخلی هر اجتماع و یال‌های مشکی بیانگر یال‌های بین جوامع هستند. کوچکی یا بزرگی رئوس رسم شده نشان دهندهٔ درجهٔ آن رأس است و همچنین قطر هر یال نشان دهندهٔ بزرگی وزن آن یال است. لازم به ذکر است اسامی نوشته شده بر روی رئوس، اسامی خلاصه شده کشورهای جدول ۱ است. همان طور که در شکل ۴ نشان داده شده است کشورهای گروه زرد رنگ شامل ایتالیا، ایالات متحده آمریکا، شیلی، فرانسه، انگلستان، اتریش ، دانمارک، سوئد، هلند، کانادا، ایرلند، پرتغال، نروژ، سوئیس، بلژیک، آرژانتین، آلمان، اسرائیل، اسپانیا، پرو، مکزیک، برزیل و فنلاند هستند و همچنین



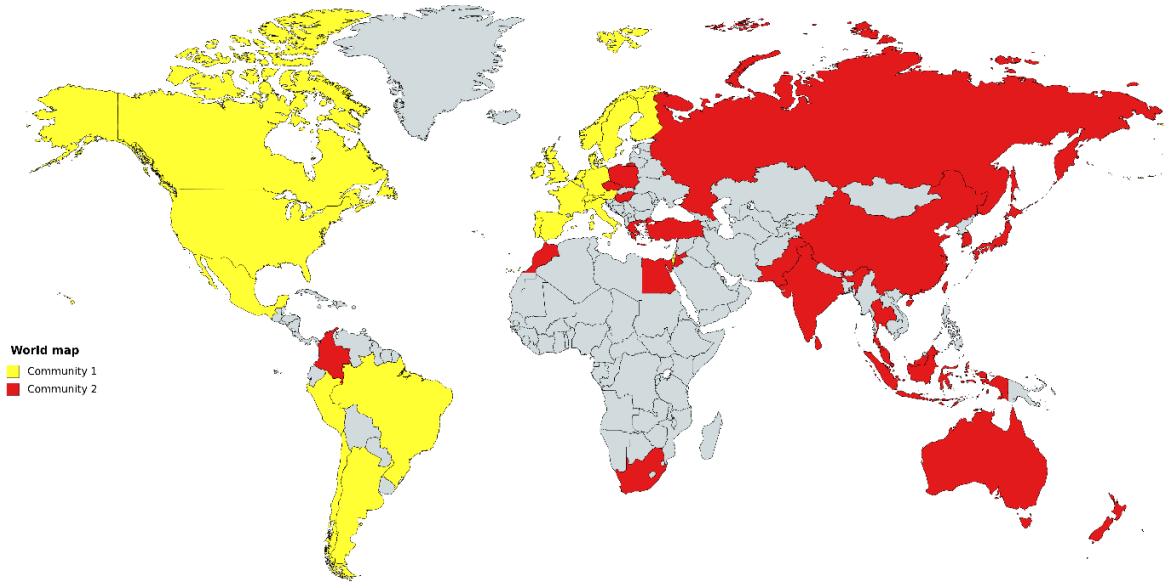
شکل ۳. مقادیر ماتریس همبستگی برای داده‌ها از ۱۹۹۵ تا ۲۰۱۹ بعد از اعمال فیلتر آستانه.



شکل ۴. شبکهٔ اقتصادی مربوط به ۴۸ کشور جدول ۱ در بازهٔ زمانی ۱۹۹۵ تا ۲۰۱۹.

کشورهای گروه قرمز رنگ شامل اردن، استرالیا، روسیه، پاکستان، مراکش، تایوان، مالزی، کره، مصر، هند، ژاپن، اندونزی، آفریقای جنوبی، نیوزیلند، تایلند، ترکیه، سریلانکا، یونان، چین، مجارستان، جمهوری چک، لهستان، کلمبیا، هنگ کنگ و سنگاپور هستند. برای درک بهتر نتایج به دست آمده از مدل استاندارد، این نتیجه را بر روی نقشهٔ جهانی کشورها در شکل ۵ نشان داده‌ایم. رنگ‌های اختصاص داده شده در این شکل مطابق با شکل ۴ است. مهم‌ترین ویژگی‌ای که از شکل ۵ استنباط می‌شود تقسیم کشورها به دو دستهٔ تقریبی شرقی و غربی از لحاظ موقعیت جغرافیایی است.

در بخش بعدی قصد داریم از رهیافت متفاوتی برای فیلتر کردن عناصر ماتریس همبستگی استفاده کنیم سپس از الگوریتم معرفی شده در بخش ۳ که برای شبکه‌های وزن‌دار و علامت‌دار شرح دادیم، برای تشکیل ساختار جوامع شبکه‌ها استفاده می‌کنیم.



شکل ۵. نمایش جوامع استخراج شده از مدل استاندارد در نقشهٔ جغرافیایی کشورها.

۴.۴. رهیافت پیشنهادی: نظریهٔ ماتریس‌های تصادفی

در این بخش با استفاده از نظریه ماتریس‌های تصادفی ابتدا به فیلتر عناصر ماتریس همبستگی می‌پردازیم و سپس با در نظر گرفتن پنجره‌های زمانی مختلف و مناسب برای تشکیل ساختار جوامع شبکه و البته بهره گرفتن از الگوریتم بیشینه‌سازی معرفی شده در بخش ۳، نگاهی دقیق‌تر به داده‌های اقتصادی کشورهای معرفی شده خواهیم داشت.

ایدهٔ این روش حذف نوفه بر اساس جابه‌جا کردن ویژه مقادیر مرتبط با نوفه است و به نام تمیزسازی ویژه مقداری نیز شناخته می‌شود [۳۵]. اولین قدم برای شروع استفاده از این روش تصمیم‌گیری راجع به این است که کدام ویژه مقادیر و ویژه بردارها مرتبط با نوفه هستند. این مسئله یک موضوع بسیار غیر بدیهی است که طبق تحقیقات قبلی متأسفانه هنوز جواب مشخصی ندارد و مرز دقیقی بین ویژه مقادیر مربوط به نوفه و ویژه مقادیر مربوط به ساختار اصلی نمی‌توان یافت. اما یک راه مرسوم برای این کار استفاده از روش زیر است:

یک ماتریس همبستگی تشکیل شده از N سری زمانی به طول T (مادامی که حد $N \to +\infty$ و $T \to +\infty$ با شرط $1 < \frac{T}{N} < +\infty$ برقرار باشد) توزیع خیلی مشخصی از ویژه مقادیرش دارد که به اسم توزیع مرچنکو- پاستور[1] یا سنگوپتا- میترا[2] [۳۵ و ۳۶] شناخته می‌شود و می‌توان آن را در رابطهٔ ۲۰ مشاهده کرد:

$$\rho(\lambda) = \begin{cases} \frac{T}{N} \frac{\sqrt{(\lambda_{Max} - \lambda)(\lambda - \lambda_{Min})}}{2\pi\sigma^2_0}, & \lambda_{Max} > \lambda > \lambda_{Min}, \\ & \lambda < \lambda_{Min} \text{ or } \lambda > \lambda_{Max} \end{cases} \quad (۲۰)$$

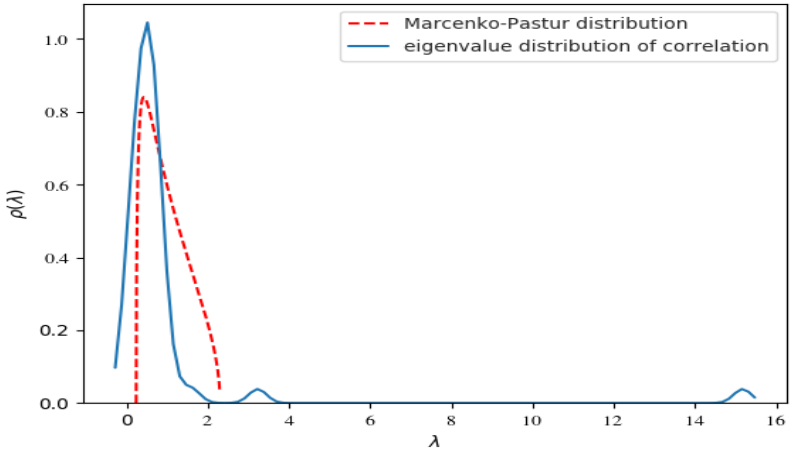
که در آن λ_{Max} و λ_{Min} از رابطهٔ زیر به دست می‌آیند:

$$\lambda_{Max,Min} = \sigma^2 (1 \pm \sqrt{N/T})^2, \quad (۲۱)$$

در روابط بالا برای بهنجار شدن ماتریس همبستگی $\sigma^2 = 1$ قرار می‌دهیم. آن قسمت از ماتریس همبستگی که مربوط به ویژه مقادیر بین $[\lambda_{Min}, \lambda_{Max}]$ است را می‌توانیم به عنوان قسمت نوفه در نظر بگیریم. بنابراین می‌توان گفت هر ویژه مقداری بزرگ‌تر از λ_{Max} را که توسط توزیع مرچنکو-پاستور پیشنهاد شده است، می‌توان به عنوان قسمت با معنی‌ای از ساختار داده در نظر گرفت و در نتیجه طبق این موضوع

۱. Marcenko-Pastur

۲. Sengupta-Mitra



شکل ۶. توزیع مارچنکو- پاستور که به رنگ قرمز ترسیم شده است و توزیع ویژه مقادیر ماتریس همبستگی برای داده‌های مدل استاندارد بخش ۴.۳ پس از اعمال لگاریتم برگشتی، که به رنگ آبی ترسیم شده است. باید اشاره کرد، بزرگ‌ترین ویژه مقدار با معنی با فاصلهٔ بسیار زیاد از بقیهٔ ویژه مقادیر قرار گرفته است.

ماتریس همبستگی C را می‌توان به دو ماتریس طبق رابطهٔ زیر تجزیه کرد:

$$C = C^{(r)} + C^{(s)}$$

که در آن قسمت نوفهٔ ماتریس همبستگی به شکل زیر است:

$$C^{(r)} \equiv \sum_{\lambda_i \le \lambda_{Max}} \lambda_i |v_i >< v_i|,$$

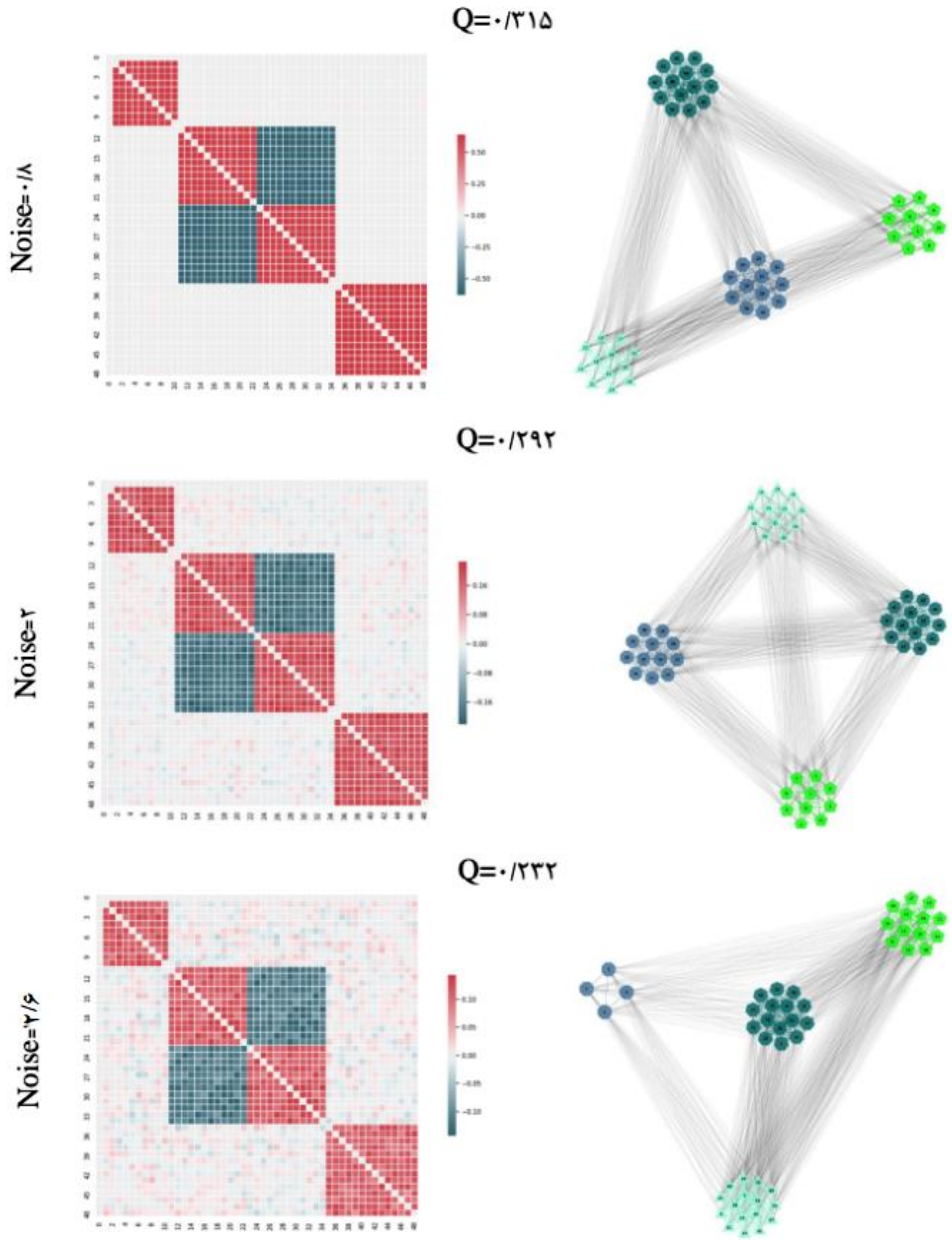
این ماتریس از ویژه مقادیر $\{\lambda_i\}$ که کوچک‌تر یا مساوی با λ_{Max} هستند تشکیل شده است (همچنین می‌توان ویژه مقادیر کوچک‌تر از λ_{Min} را نیز در نظر گرفت). پس می‌توان گفت ماتریس همبستگی مربوط به ساختار اصلی داده‌ها به صورت $C^{(s)} = C - C^{(r)}$ است.

اختلاف میان طیف اصلی ویژه مقادیر ماتریس همبستگی از مدل نظریهٔ ماتریس تصادفی، راهی مناسب را برای ما ایجاد می‌کند تا بتوانیم نوفه را از داده‌های اصلی حذف کنیم. برای واضح شدن موضوع می‌توان به شکل ۶ دقت کرد. در این نمودار توزیع ویژه مقادیر ماتریس همبستگی که از ۴۸ کشور جدول ۱ تشکیل می‌شود ($N = ۴۸$)، به همراه توزیع مرچنکو- پاستور مورد انتظارش با همان مقادیر N رسم شده است و همان طور که در مطالعات دیگر [۳۷] نیز دیده شده است یکی از خواص توزیع اصلی ویژه مقادیر رسم شده، فاصله داشتن بزرگ‌ترین ویژه مقدار از اکثریت دیگر توزیع است. در مواردی که در مورد یک بازار مالی داخلی در حال صحبت هستیم (به طور مثال بازار سهام یک کشور به تنهایی) می‌توانیم این بزرگ‌ترین ویژه مقدار را به عنوان مُد جهانی بازار در نظر بگیریم که یک تاثیر از بازار جهانی کلی است.

قبل از آن که رهیافت خود را بر روی ماتریس همبستگی اصلی اعمال کنیم قصد داریم این رهیافت را به صورت آزمایشی بررسی کنیم تا مشخص شود آیا این الگوریتم توانایی تشخیص مجموعه‌های مرتبط از سری‌های زمانی یک سری داده ساختگی را دارد یا خیر. به همین منظور شروع به ساخت سری‌های زمانی با جوامع دلخواه و مشخص می‌کنیم سپس در هر مرحله نوفهٔ این سری‌های زمانی را بالا خواهیم برد تا از عملکرد درست الگوریتم استفاده شده، اطمینان حاصل کنیم. باید اشاره کرد که روند طی شده بسیار شبیه به مدل‌های عاملی[۱] استفاده شده در تحلیل‌های مالی است [۳۹ و ۴۰].

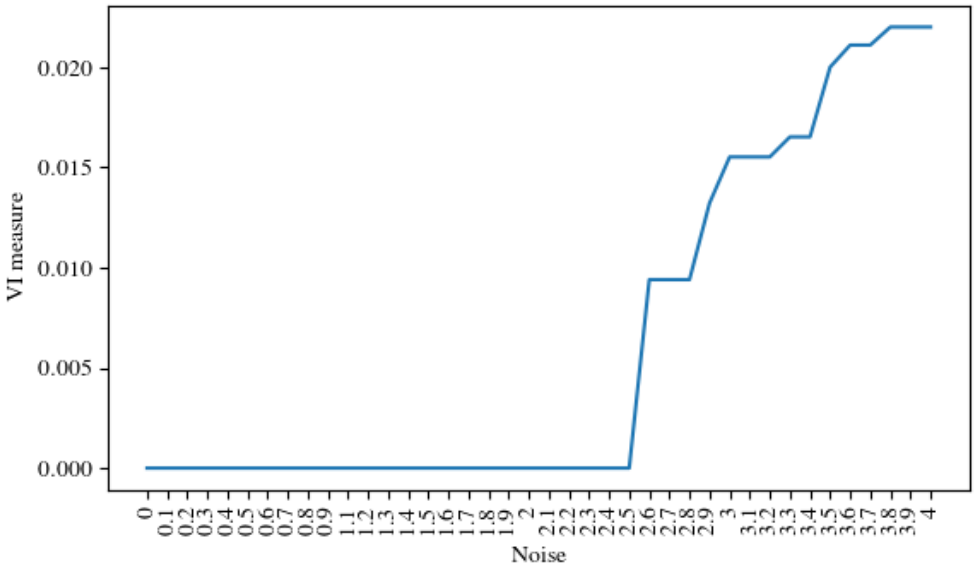
ما برای ساخت داده‌های آزمایش از جوامع با اندازهٔ متفاوت و البته میزان نوفهٔ متفاوت استفاده کردیم تا درستی رهیافت پیشنهادی را بسنجیم. رهیافت پیشنهادی استفاده شدهٔ ما در اینجا فیلتر داده‌ها با استفاده از نظریهٔ ماتریس‌های اتفاقی معرفی شده در بخش ۴.۴ و خوشه‌بندی با استفاده از

۱. Factor models



شکل ۷. بررسی درستی عملکرد رهیافت پیشنهادی بر روی داده‌های ساختگی. این داده‌ها، سری‌های زمانی همراه با مقدار ضریب نوفهٔ متفاوت هستند و در سری داده‌ها از ۴۸ سری زمانی، به ترتیب ۱۰، ۱۲، ۱۲ و ۱۴ عضو آن با هم تشکیل خوشه می‌دهند. برای هر سری آزمایشی نتیجه مقدار پیمانگی بعد از تقسیم‌بندی نیز در بالای آن ذکر شده است.

بیشینه‌سازی پیمانگی وزن‌دار شده بخش۳ است. در شکل ۷ سه سری دادهٔ آزمون مطابق با نمایش مقدار ماتریس همبستگی آنها در سمت چپ شکل، به ترتیب با مقدار ضریب نوفه ۰/۸، ۲ و ۲/۶ ساخته‌ایم. این داده‌های آزمون شامل ۴ جامعه به اندازه‌های ۱۰، ۱۲، ۱۲ و ۱۴ عضو (و مجموع $N=۴۸$ عضو) برای شباهت با داده‌های اصلی هستند. این جوامع همبستگی کامل داخلی دارند و با جوامع دیگر غیر همبسته‌اند. البته برای در نظر گرفتن کلی‌ترین حالت و نشان دادن درستی عملکرد فرمول‌بندی بخش ۳، دو جامعهٔ ۱۲ عضوی پادهمبستگی کامل برای اعضای بین جوامع دارند. در نظر گرفتن این حالت که در فرمول بندی اصلی کاملا به خطا منجر می‌شود برای بررسی رهیافت بسیار اساسی است.



شکل ۸. تغییر اطلاعات خوشه یافته شده نسبت به خوشهٔ اصلی برای ضریب نوفهٔ ۰ تا ۴ در مدل آزمون.

همچنین مقدار پیمانگی بعد از خوشه‌بندی، نیز برای مقایسهٔ نزدیکی سیستم به سیستم اتفاقی در بالای هر مجموعه دادهٔ آزمون ذکر شده است. مشاهده می‌شود که رهیافت ما تا مقدار ضریب نوفه ۲/۶ به درستی جوامع را تشخیص داده است و در این مقدار نوفه برای اولین بار از تشخیص کامل کوچک‌ترین اندازهٔ جامعه (یعنی جامعه با اندازهٔ ۱۰) ناتوان بوده است و تنها ۴ عضو از ۱۰ عضو این جامعه را به درستی تشخیص داده است. با توجه به این که مقدار ضریب نوفه ۲/۶ می‌تواند بسیار بزرگ باشد در نتیجه می‌توان برداشت کرد که رهیافت معرفی شده تا حد خوبی نسبت به نوفه و داده‌های اتفاقی سری زمانی، مقاوم[1] است. برای این که مقاومت رهیافت پیشنهادی نسبت به نوفه را بتوانیم بهتر نشان دهیم از معیاری به اسم تغییر اطلاعات استفاده خواهیم کرد [۴۰].

به همین منظور دو گروه‌بندی متفاوت یک شبکه را به شکل‌های $Y=\{Y_1,Y_2,\ldots\}$ و $X=\{X_1,X_2,\ldots\}$ در نظر بگیرید که در آنها تعداد اعضای شبکه ثابت است و با n نشان داده می‌شود. با توجه به موارد گفته شده کمیت تغییر اطلاعات را می‌توان به شکل زیر تعریف کرد:

$$VI(X;Y)=-\frac{1}{\log(n)}\sum_{ij} r_{ij}\left[\log\left(r_{ij}/p_i\right)+\log\left(r_{ij}/q_j\right)\right], \quad (۲۲)$$

در اینجا کمیت‌های $p_i=\frac{|X_i|}{n}$ ، $q_j=\frac{|Y_J|}{n}$ و $r_{ij}=\frac{|X_i\cap Y_J|}{n}$ هستند. در شکل ۸ این تغییر اطلاعات بهنجار شده برای بازهٔ ضریب نوفهٔ ۰ تا ۴ رسم شده است. همان طور که مشاهده می‌شود این رهیافت تا حد بالایی از نوفه مقاوم است.

لازم است تأکید شود در صورتی که تعداد اعضای جوامع موجود نسبت به یکدیگر متفاوت باشد در آن صورت، در حد نوفهٔ کمتری توانایی شناسایی جامعه با تعداد اعضای کمتر، از بین خواهد رفت. در ادامه قصد داریم تا با معرفی پنجرهٔ زمانی مناسب، این رهیافت پیشنهادی را بر روی داده‌های اصلی اعمال کنیم.

۵.۴. اعمال رهیافت پیشنهادی بر روی داده‌های مورد بررسی

در این بخش بعد از ذکر نتایج رهیافت پیشنهادی بر روی داده‌های اصلی بدون پنجرهٔ زمانی، به معرفی پنجرهٔ زمانی

۱. Robust

شکل ۹. به ترتیب مقدار پیمانگی برای پنجره‌های زمانی ۱ ساله در بازهٔ زمانی ۱۹۹۵ تا ۲۰۱۹ با رنگ آبی و معیار تغییرات اطلاعات برای همین بازهٔ زمانی با رنگ نارنجی.

مناسب برای بررسی جزییات بیشتر از داده‌های اصلی می‌پردازیم و سپس با استفاده از نظریهٔ ماتریس‌های اتفاقی به فیلتر کردن داده‌های اصلی پرداخته تا برای اعمال مدل بیشینه‌سازی وزن دار معرفی شده در بخش ۳ آماده شوند.

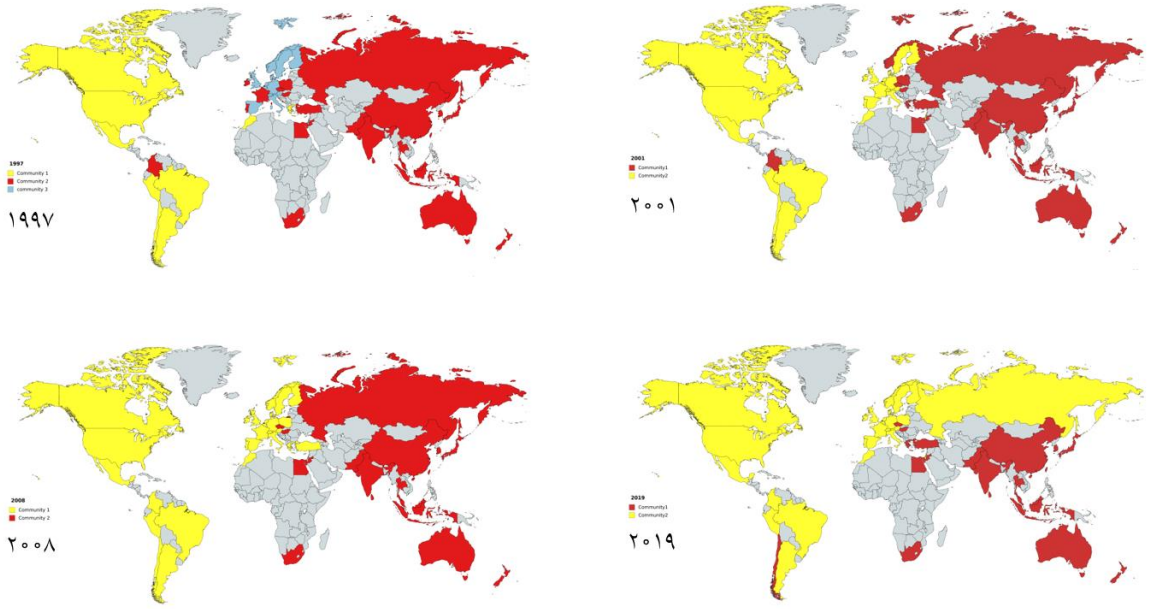
۱.۵.۴ بررسی جوامع بدون پنجرهٔ زمانی

در ابتدا قصد داریم نتایج بخش ۳.۴ را با رهیافت پیشنهادی، مقایسه کنیم. زمانی که رهیافت پیشنهادی را برروی داده‌ها بدون پنجرهٔ زمانی اعمال می‌کنیم نتایج به دست آمده کاملا مطابق با رهیافت استاندارد بخش ۲ خواهد بود. علت آن را می‌توان عدم وجود همبستگی‌های منفی در زمان بررسی داده‌ها بدون پنجرهٔ زمانی، دانست. البته باید اشاره کرد دو نوع فیلتر معرفی شده نیز (حد آستانه و نظریهٔ ماتریس تصادفی) در طول زیاد داده رفتاری مشابه دارند. به همین دلیل از تکرار نتایج بخش ۳.۴ خودداری خواهیم کرد.

۲.۵.۴ بررسی جوامع در پنجرهٔ زمانی ۱ ساله

در ابتدا به سراغ تشکیل ماتریس همبستگی و فیلتر قسمت اتفاقی آن خواهیم رفت. باید اشاره کرد که در نظر گرفتن پنجرهٔ زمانی ۱ ساله (شامل ۲۶۰ روز معاملاتی) منجر به تشکیل ۲۵ پنجرهٔ زمانی برای کل داده‌ها می‌شود. پس از آن که عناصر ماتریس همبستگی را با توجه به توزیع ویژه مقادیر ذکر شده فیلتر کردیم، می‌توانیم ساختار جوامع مربوط به این پنجره‌های ۱ ساله را به شکل پیوسته مشاهده کنیم و در هر پنجره با اندازه‌گیری مقدار پیمانگی نهایی Q به قیاس بپردازیم. حال می‌توانیم به بررسی تغییرات مقدار پیمانگی در طی کل بازه زمانی و با پنجرهٔ زمانی ۱ ساله بپردازیم. در شکل ۹ این مقدار رسم شده است. همان طور که مشاهده می‌شود در پنجرهٔ زمانی ۲۰۰۸ این مقدار به کمترین حد خود در طی کل یازه یعنی $Q = ۰.۰۳۱$ رسیده است. اگر نگاهی دوباره به رابطهٔ اصلی پیمانگی یعنی رابطهٔ ۱ بیاندازیم متوجه خواهیم شد که کم بودن این مقدار به معنی میل کردن شبکه به حالت اتفاقی خود خواهد بود. همان طور که مشاهده می‌شود نتایج ما در این بازهٔ زمانی که بحران سال ۲۰۰۸ بوده است نشان دهندهٔ رفتار اتفاقی بازارهای سرمایه به نسبت بقیه پنجره‌های زمانی است.

همچنین در نمودار ۹ می‌توان میزان تغییرات هر خوشه نسبت به خوشهٔ بعدی را با پنجرهٔ زمانی ۱ ساله و رنگ نارنجی مشاهده کرد. این تغییر اطلاعات در واقع همان رابطهٔ ۲۲



شکل ۱۰. ساختار جوامع مربوط به پنجرهٔ زمانی ۱ ساله برای سال‌های ۱۹۹۷، ۲۰۰۱، ۲۰۰۸ و ۲۰۱۹. در این شکل‌ها رنگ زرد و قرمز و آبی بیانگر ساختار جوامع پیدا شده است.

است که برای آزمون داده‌ها با نوفهٔ متفاوت نیز استفاده شد. همان طور که از نمودار مشاهده می‌شود تغییرات خوشه‌بندی در سال ۲۰۱۱ نسبت به خوشهٔ قبلی عددی نزدیک به صفر بوده است. تغییرات بسیار اندک خوشه بعد از یک بحران مالی نیز قابل توجه است.

در ادامه، ۴ نمونه از این ساختار یافته شده را برروی نقشهٔ جهان نشان داده‌ایم. این ۴ نمونه که در شکل ۱۰ نشان داده شده‌اند، جوامع یافته شده در سال‌های ۱۹۹۷، ۲۰۰۱، ۲۰۰۸ و ۲۰۱۹ هستند. همچنان بارزترین ویژگی که از جوامع یافته شده استخراج می‌شود ارتباط جغرافیایی قوی بین کشورهای همسایه است. همچنین مشاهده می‌شود به طور مثال در مدل استخراجی سال ۱۹۹۷ جوامع یافته شده ۳ بوده است.

۳.۵.۴. بررسی جوامع در پنجرهٔ زمانی ۲ ساله

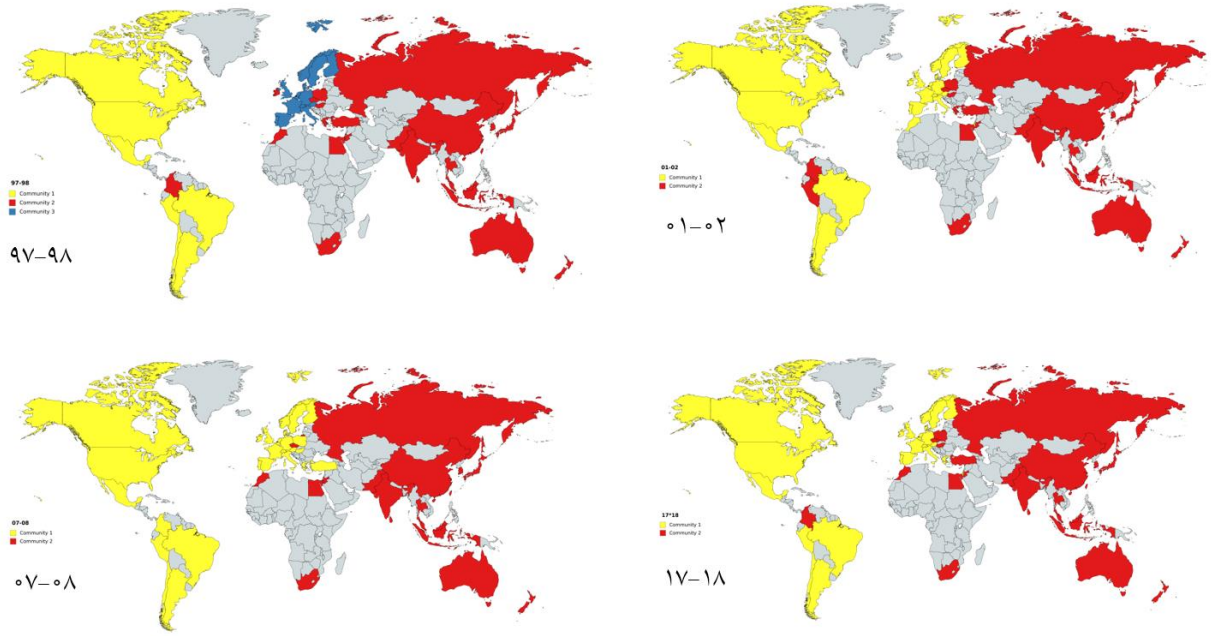
در این قسمت دوباره به فیلتر داده‌ها با استفاده از نظریهٔ ماتریس‌های اتفاقی می‌پردازیم سپس با استفاده از الگوریتم بیشینه‌سازی پیمانگی وزن‌دار به استخراج ساختار جوامع می‌پردازیم. در ابتدا قصد داریم تا مقدار پیمانگی یعنی Q را مانند شکل ۹ برای این پنجرهٔ زمانی بررسی کنیم. در شکل ۱۲ میزان مقدار پیمانگی و تغییر اطلاعات در طی بازهٔ زمانی ۱۹۹۵ تا ۲۰۱۸ برای پنجرهٔ ۲ ساله (۵۲۰ روز معاملاتی) مشاهده می‌شود. نکتهٔ جالب توجه ۳ دورهٔ کاملا مشخص با کمترین میزان پیمانگی در این شکل است. به ترتیب دوره‌های ۲۰۰۸–۲۰۰۹ به دلیل رخ دادن بحران مالی جهانی، ۲۰۱۵–۲۰۱۶ به دلیل رخ دادن بحران در بازارهای مالی چین و تاثیر چین به عنوان رهبر مشخص جوامع قرمز رنگ و ۱۹۹۷–۱۹۹۸ به دلیل رخ دادن بحران اقتصادی در بازارهای جنوب شرقی آسیا، کمترین مقدار پیمانگی را دارد. به طور کلی بازهٔ زمانی ۲ ساله به دلیل اعتبار بیشتر در شرط فیلتر ماتریس‌های تصادفی (حد $N \to +\infty$ و $T \to +\infty$ با شرط برقراری $1 < \frac{T}{N} < +\infty$) و البته داده‌های بیشتر، دید کلی بهتری را به نسبت پنجرهٔ زمانی ۱ ساله، برای تحلیل جوامع در طی این بازهٔ زمانی، برای ما فراهم می‌کند. در شکل ۱۱ همچنان تغییر بسیار کم خوشه‌ها بعد از هر بحران مالی نکته قابل توجهی است.

شکل ۱۱. به ترتیب مقدار پیمانگی برای پنجره‌های زمانی ۲ ساله در بازهٔ زمانی ۱۹۹۵ تا ۲۰۱۸ با رنگ آبی و معیار تغییرات اطلاعات برای همین بازهٔ زمانی با رنگ نارنجی.

در ادامه نیز جوامع استخراج شده در ۴ نمونه از این ۱۲ پنجرهٔ زمانی تشکیل شده با استفاده از بازهٔ ۲ ساله، بر روی نقشهٔ جغرافیایی کشورها در شکل ۱۲ نشان داده شده است. همان طور که در شکل ۱۲ دیده می‌شود همچنان دودستگی کلی‌ای به شکل شرقی و غربی در جوامع یافته شده مشاهده می‌شود که با انحراف کمی تایید کنندهٔ شکل کلی جوامع یافته شده با استفاده از پنجرهٔ زمانی ۱ ساله است.

۵. جمع‌بندی و نتیجه‌گیری

کاربرد روش‌های آماری همواره برای توصیف شبکه‌های اقتصادی مفید بوده است. این ابزارها ما را قادر به بررسی و استخراج اطلاعاتی از یک شبکهٔ اقتصادی پیچیده می‌کند. به همین منظور در این مقاله به توصیف یکی از این ابزارهای بزرگ مقیاس به نام خوشه‌یابی یا تشخیص جامعه، پرداختیم. در ادامه یکی از الگوریتم‌های خاص این دسته از ابزارها را به نام بیشینه‌سازی پیمانگی، معرفی کردیم. با استفاده از استخراج ماتریس مجاورت شبکه از همبستگی سری‌های زمانی بازارهای مالی ۴۸ کشور مورد بررسی و استفاده از حد آستانه مناسب، خوشگی مربوطه را در بازه زمانی ۱۹۹۵ تا ۲۰۱۹ و با استفاده از الگوریتم استاندارد بیشینه‌سازی پیمانگی، استخراج کردیم. در این قسمت به عنوان اولین نتیجه مشاهده شد این خوشگی به شدت تحت تاثیر موقعیت جغرافیایی کشورها بوده است. اما به دلیل حذف ارتباط‌های منفی شبکه در الگوریتم استاندارد بیشینه‌سازی پیمانگی و البته مناسبت نبودن فیلتر داده‌ها با استفاده از حد آستانه، در ادامه به توسعه بیشینه‌سازی پیمانگی برای شبکه‌های وزن‌دار و علامت‌دار و البته فیلتر داده‌ها با استفاده از نظریهٔ ماتریس‌های اتفاقی پرداختیم. برای توصیف مناسب این روش، داده‌ها را در دو پنجرهٔ زمانی ۱ و ۲ ساله بررسی کردیم و همان طور که دیدیم با الگوریتم توسعه‌یافته نیز خوشگی به شدت تحت تاثیر موقعیت جغرافیایی شبکه است و به طور کلی عموماً این بخش‌بندی به دو قسمت شرقی و غربی، از نظر جغرافیایی، تقسیم‌بندی می‌شود. با بررسی مقدار پیمانگی در طی پنجره‌های زمانی یاد شده، مشاهده کردیم بازارهای مالی به ترتیب، کاتوره‌ای‌ترین رفتار خود را در بازهٔ بحران جهانی



شکل ۱۲. ساختار جوامع مربوط به پنجرهٔ زمانی ۲ ساله برای سال‌های ۱۹۹۷–۹۸، ۲۰۰۱–۰۲، ۲۰۰۷–۰۸ و ۲۰۱۷–۱۸. در این شکل‌ها رنگ زرد، قرمز و آبی بیانگر ساختار جوامع پیدا شده است.

اقتصاد ۲۰۰۸، بحران بازارهای مالی چین در ۲۰۱۵ و بحران اقتصادی آسیای جنوب شرقی ۱۹۹۷ نشان داده‌اند. همان طور که گفته شد بیشترین تمرکز این پژوهش بر روی همبستگی خطی بوده و از آنجایی که ارتباط بین داده‌های اقتصادی لزوماً خطی نبوده، یکی از پیشنهادها برای ادامهٔ پژوهش بررسی رهیافت‌های غیر خطی جهت تشکیل شبکهٔ اقتصادی از سری‌های زمانی ارزش بازار و استخراج جوامع غیر خطی در بازه‌های زمانی مناسب است. الگوریتم بیشینه‌سازی پیمانگی گرچه ویژگی‌های مثبتی همچون عدم نیاز به ناظر و دقت بالا در تشکیل جوامع دارد اما این الگوریتم هیچ اطلاعاتی راجع به اتصال‌های بین خوشه‌ها در اختیار ما قرار نمی‌دهد؛ به طوری که جابه‌جا کردن این اتصال‌های بین گروهی تاثیری در نتیجه نهایی نخواهد داشت. در نتیجه یکی از پیشنهادها برای ادامهٔ این پژوهش، بسط یا جایگزینی الگوریتم بیشینه‌سازی پیمانگی با الگوریتمی است که به اتصال‌های بین جوامع نیز حساس باشد. به عنوان یک کاربرد مفید می‌توان از این الگوریتم در راستای بررسی ساختار جوامع بازار بورس داخلی و پیشنهاد کردن معیار برای سرمایه‌گذاری در چند بخش مختلف از این بازار در راستای بالا بردن بهره‌وری سرمایه‌گذاری استفاده کرد. به عنوان یکی دیگر از محرک‌های موجود برای ادامهٔ پژوهش، می‌توان از تمیزسازی و استفاده از دادهٔ بازار بورس ایران برای اضافه شدن به لیست ۴۸ کشور بررسی شده و بررسی عضویت اقتصاد ایران در هر یکی از این جوامع در طی تاریخ یاد کرد.

مراجع

1. A Barabási. "*Network science*". Cambridge university press (2016).
2. J A Dunne, et al., *Proceedings of the National Academy of Sciences* **99**, 20 (2002) 12917.
3. M Boss, et al. *Quantitative finance* **4**, 6 (2004) 677.
4. M E Newman. *Nature physics* **8**, 1 (2012) 25.
5. S Fortunato. *Physics reports* **486**, 3-5 (2010) 7.
6. P J Mucha., et al., *science* **328,** 5980 (2010) 876.
7. M Girvan and M E Newman. *Proceedings of the national academy of sciences* **99**, 12 (2002) 7821.
8. A Lancichinetti, S Fortunato, and F Radicchi. *Physical Review E* **78**, 4 (2008) 046110.
9. L Danon, et al., *Journal of Statistical Mechanics: Theory and Experiment* **09** (2005) P09008.

10. A Lancichinetti, S Fortunato, and F Radicchi, *Physical Review* **80**, 1 (2009) 016118.
11. A Pothen, H D Simon, and K P. Liou, *SIAM journal on matrix analysis and applications* **11**, 3 (1990) 430.
12. B W Kernighan and S Lin. *The Bell system technical journal* **49**, 2 (1970) 291.
13. D Krackhardt and R N Stern, *Social psychology quarterly* (1988) 123.
14. B Karrer and M E Newman, *Physical Review E* **83**, 1 (2011) 016107.
15. Z Li, et al., *Physical Review E* **77**, 3 (2008) 036109.
16. M. E. Newman, and M. Girvan. *Physical Review E* **69**, 2 (2004) 026113.
17. A Medus, et al, *Physica A: Statistical Mechanics and its Applications* **358**, 2-4 (2005) 593.
18. R L Breiger, S A Boorman, and P. Arabie, *Journal of mathematical psychology* **12** (1975) 328.
19. P W Holland, K B Laskey, and S Leinhardt, *Social networks* **5**, 2 (1983) 109.
20. F Bernardeau, et al., *Physics Reports* **367**, 1-3 (2002) 1.
21. A Vafaei Sadr, and S M S Movahed, *Monthly Notices of the Royal Astronomical Society* **503,** 1 (2021) 815.
22. R K Pathria, and P D Beale, "*Statistical Mechanics”*, edition. (2011).
23. D Enke, M Grauer, and N Mehdiyev, *Procedia Computer Scienc*e **6** (2011) 201.
24. C Y Chiu, et al., *Expert Systems with Applications* **36**, 3 (2009) 4558.
25. K J Kim and H Ahn, *Expert systems with applications* **34**, 2 (2008) 1200.
26. A J Lee, et al., "*An Effective Clustering Approach to Stock Market Prediction*" In PACIS. (2010.)
27. M S Prieto and A R Allen, *Image and Vision Computing* **27**, 6 (2009) 673.
28. M C Münnix, et al., *Scientific reports* **2** (2012) 644.
29. M. E. Newman, *Proceedings of the national academy of sciences* **103**, 23 (2006) 8577.
30. F R Chung and F C Graham, No. 92. American Mathematical Soc. (1997).
31. M Fiedler, *Czechoslovak mathematical journal* **23**, 2 (1973) 298.
32. S Gómez, P Jensen, and A Arenas, *Physical Review E* **80**, 1 (2009) 016114.
33. P Ferreira, A Dionísio, and S M S Movahed, *Physica A: Statistical Mechanics and its Applications* **486** (2017) 730.
34. R A Fisher, Biometrika **10**, 4 (1915) 507.
35. L Laloux, et al., *Physical review letters* **83**, 7 (1999) 1467.
36. V. Plerou, et al., *Physical Review Letters* **83**, 7 (1999) 147.
37. J P Bouchaud, and M Potters, Cambridge university press, (2003).
38. M Tumminello, F Lillo, and R N Mantegna, *AIP Conference Proceedings. American Institute of Physic*s **965** (2007) 1
39. E F Fama., et al. *Financial Analysts Journal* **49**, 1 (1993) 37.
40. M Marina, *Journal of multivariate analysis* **98**, 5 (2007) 873.