

کاربرد روش‌های خوشه‌بندی در ترسیم

نقشه‌های علم: مورد کاوی نقشه علم

مدیریت شهری

محمد ابوبی اردکان

دکترای مدیریت

استادیار دانشکده مدیریت دانشگاه تهران

حسن عابدی جعفری

دکترای مدیریت

استادیار دانشکده مدیریت دانشگاه تهران

فتاح آقازاده دهده*

کارشناس ارشد مدیریت شهری دانشگاه تهران

اطلاعات
علوم و فناوری

دریافت: ۱۳۸۸/۰۸/۰۷ | پذیرش: ۱۳۸۸/۱۰/۰۷ | مقاله برای اصلاح به مدت ۱۵ روز نزد پدیدآوران بوده است.

چکیده: هدف از نگارش مقاله حاضر، آشنایی مقدماتی با خوشه‌بندی اطلاعات و نشان دادن کاربرد روش‌های خوشه‌بندی در ترسیم نقشه‌های علم می‌باشد. به این منظور انواع رویکردها به دسته‌بندی و خوشه‌بندی اطلاعات به صورت مختصر مورد بررسی قرار می‌گیرند و با بررسی پژوهش‌های مشابه در این شاخه علمی، کاربرد این روش‌ها در فرایند مصورسازی اطلاعات مفهومی و ترسیم نقشه‌های علم معرفی می‌شود. در انتها با اجرای الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی که نوعی از الگوریتم‌های خوشه‌بندی بر مبنای روش ارتباط کامل می‌باشد، نقشه علم مدیریت شهری به‌عنوان یکی از شاخه‌های علمی بین رشته‌ای و نوظهور، مورد تحلیل و بررسی قرار می‌گیرد.

کلیدواژه‌ها: خوشه‌بندی اطلاعات؛ علم‌سنجی؛ نقشه‌های علم؛ نقشه علم مدیریت شهری

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شبا (چاپی) ۵۲۰۶-۱۷۳۵
شبا (الکترونیکی) ۵۵۸۳-۲۰۰۸
نمایه در LISA و SCOPUS
<http://jlist.irandoc.ac.ir>
دوره ۲۵ | شماره ۳ | صص ۳۴۷-۳۷۱
بهار ۱۳۸۹

نوع مقاله: علمی پژوهشی

* پدیدآور رابط fattah.aghazadeh@gmail.com

۱. مقدمه

خوشه‌بندی به معنای دسته‌بندی اعضای مجموعه‌ها بدون نظارت و دخالت است. در این روش، خوشه‌ها یا دسته‌ها از قبل تعیین شده نیستند و به عبارت دیگر، برچسب خوشه‌ها در دسترس نیست (Jain, Murty and Flynn 1999). خوشه‌بندی، به یافتن ساختاری در درون یک مجموعه از داده‌های بدون برچسب اطلاق می‌شود؛ خوشه به مجموعه‌ای از داده‌ها گفته می‌شود که به هم شباهت داشته باشند. در خوشه‌بندی سعی می‌شود داده‌ها به خوشه‌هایی تقسیم شوند که شباهت بین داده‌های درون هر خوشه، حداکثر و شباهت بین داده‌های درون خوشه‌های متفاوت، حداقل شود (حاج‌احمدی ۱۳۸۵).

طبقه‌بندی و خوشه‌بندی با هم متفاوت‌اند. در طبقه‌بندی، هر داده به یک طبقه یا کلاس از پیش مشخص شده تخصیص می‌یابد؛ ولی در خوشه‌بندی، هیچ اطلاعی از کلاس‌های موجود در درون داده‌ها وجود ندارد و به عبارتی، خود خوشه‌ها نیز از داده‌ها استخراج می‌شوند (Engelbrecht and Salman 2007). در نتیجه پس از انجام خوشه‌بندی، یک فرد خبره باید خوشه‌های ایجادشده را تفسیر کند و در بعضی مواقع لازم است که پس از بررسی خوشه‌ها، بعضی از پارامترهایی که در خوشه‌بندی در نظر گرفته شده‌اند ولی بی‌ربط هستند یا اهمیت چندانی ندارند، حذف شوند و خوشه‌بندی از اول صورت گیرد (Edelstein 1999).

تحلیل خوشه‌ها در حقیقت سازماندهی مجموعه‌ای از نمونه‌ها به خوشه‌ها بر پایه تشابهات می‌باشد - یعنی این که نمونه‌هایی که در یک خوشه قرار دارند، ویژگی‌های مشابه‌تری نسبت به هم دارند. وجود شیوه‌های متعدد برای نمایش داده، اندازه‌گیری تشابه بین عناصر داده‌ای و گروه‌بندی آن‌ها باعث شده است که شیوه‌های خوشه‌بندی مختلفی داشته باشیم (Flynn, Jain and Murty 1999). اگرچه امروزه تمایل بسیار زیادی برای به کارگیری روش‌های مختلف خوشه‌بندی در زمینه‌های مختلف از قبیل شناسایی الگوها^۱ (Anderberg 1973)، پردازش تصویر (Jain and Flynn 1996)، و بازیابی اطلاعات (Rasmussen 1992) وجود دارد، این روش‌ها در زمینه‌های دیگر، از سابقه طولانی‌تری برخوردار هستند. خوشه‌بندی در بسیاری از شاخه‌های علوم کاربرد دارد، که از جمله می‌توان به موارد زیر اشاره داشت (Duda, Hart and Stork 2000, 69):

^۱ pattern recognition

- ◇ بازاریابی: دسته‌بندی مشتری‌ها به دسته‌هایی بر حسب رفتارها و نیازهای آن‌ها از طریق مجموعه‌ی زیادی از ویژگی‌ها و آخرین خریدهای آن‌ها؛
- ◇ زیست‌شناسی: دسته‌بندی جانوران و گیاهان از روی ویژگی‌های آن‌ها؛
- ◇ کتابداری: دسته‌بندی کتاب‌ها و دیگر منابع اطلاعاتی؛
- ◇ نقشه‌برداری شهری: دسته‌بندی خانه‌ها بر اساس نوع و موقعیت جغرافیایی آن‌ها؛
- ◇ مطالعات زلزله‌نگاری: تشخیص مناطق حادثه‌خیز بر اساس مشاهدات قبلی؛
- ◇ وب: دسته‌بندی اسناد یا محتوای اطلاعاتی مربوط سایت‌ها یا کاربران پایگاه‌های وب؛
- ◇ داده‌کاوی: کشف اطلاعات و ساختار جدید از داده‌های موجود.

هدف از نگارش مقاله حاضر، آشنایی با کاربرد روش‌های خوشه‌بندی در ترسیم نقشه‌های علم است. بدین منظور با بررسی انواع رویکردها و روش‌های رایج در خوشه‌بندی، فوننی که در ترسیم نقشه‌های علم کاربرد بیش‌تری دارند به اختصار توضیح داده خواهند شد.

اهمیت خوشه‌بندی و ماهیت بین‌رشته‌ای آن باعث شده است که ادبیات تحقیق و نوشته‌های بسیار زیادی در خصوص خوشه‌بندی موجود باشد. کتاب‌های زیادی در این زمینه منتشر شده‌اند و مقالات مروری مفیدی در این زمینه انتشار یافته‌اند (Anderberg 1973; Hartigan 1975; Spath 1980; Backer 1995; Jain and Dubes 1998).

۲. روش‌های خوشه‌بندی

الگوریتم‌های خوشه‌بندی داده‌ها در قالب سه رویکرد کلی بررسی می‌شوند: الگوریتم‌های سلسله‌مراتبی^۱، الگوریتم‌های افزایی^۲، و الگوریتم‌های فازی^۳.

۱-۲. خوشه‌بندی سلسله‌مراتبی

در روش‌های خوشه‌بندی سلسله‌مراتبی، به خوشه‌های نهایی بر اساس میزان عمومیت آن‌ها، ساختاری سلسله‌مراتبی نسبت داده می‌شود. به ساختار سلسله‌مراتبی حاصل از روش‌های خوشه‌بندی سلسله‌مراتبی، نمودار دندان‌های^۴ گفته می‌شود (Jain, Murty and

¹ hierarchical clustering algorithms

² prtitional custering agorithms

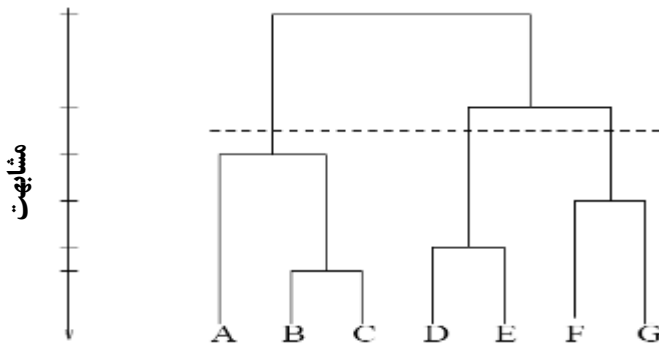
³ fzyzy clustering agorithms

⁴ Dendogram

(Flynn 1999). روش های خوشه بندی، بر اساس ساختار سلسله مراتبی تولیدی توسط آنها معمولاً به دو دسته زیر تقسیم می شوند:

بالا به پایین یا تقسیم کننده: در این روش ابتدا تمام داده ها به عنوان یک خوشه در نظر گرفته می شوند و سپس در طی یک فرایند تکراری، در هر مرحله داده هایی که شباهت کم تری به هم دارند به خوشه های مجزا شکسته می شوند و این روال تا رسیدن به خوشه هایی که دارای یک عضو هستند، ادامه پیدا می کند.

پایین به بالا یا متراکم شوند: در این روش ابتدا هر داده به عنوان خوشه ای مجزا در نظر گرفته می شود و در فرایندی تکراری، در هر مرحله خوشه هایی که شباهت بیش تری با یکدیگر دارند با هم ترکیب می شوند تا در نهایت یک خوشه حاصل شود. الگوریتم های سلسله مراتبی نسبت به الگوریتم های افزایی که در قسمت بعد آنها را معرفی می کنیم، از تنوع بیش تری برخوردارند (Jain, Murty and Flynn 1999).



شکل ۱. نمونه ای از نمودار دندانهای شکل (Jain, Murty and Flynn 1999)

از انواع الگوریتم های خوشه بندی سلسله مراتبی رایج می توان از الگوریتم های تک اتصال، میانگین اتصال^۱، میانگین اتصال گروهی^۲ و اتصال کامل^۳ نام برد. تفاوت اصلی

^۱ average-link

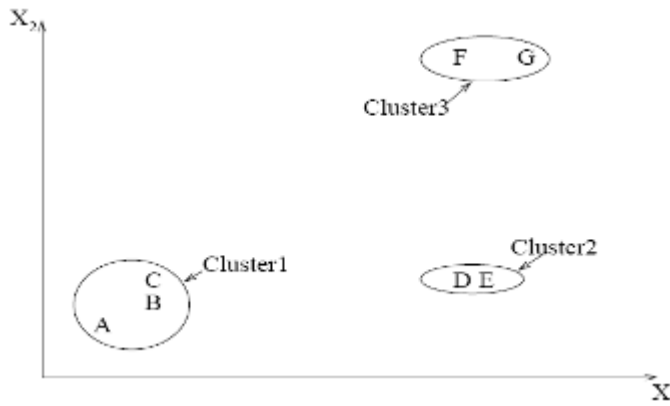
^۲ group average link

^۳ complete-link

در بین این روش‌ها به نحوه محاسبه شباهت بین خوشه‌ها مربوط می‌شود (Flynn, Jain and Murty 1999).

۲-۲. خوشه‌بندی افزایشی

آنچه که از یک الگوریتم خوشه‌بندی افزایشی به دست می‌آید، یک گروه واحد از داده به جای یک ساختار خوشه‌ای می‌باشد. مشکل الگوریتم‌های افزایشی این است که تعداد خوشه‌های خروجی چگونه انتخاب شود. روند تولید خوشه‌ها در روش‌های افزایشی بر این اساس است که یک تابع سنجش^۱ از قبل تعریف شده، فرایند را بهینه کند که محاسبه مقدار بهینه برای آن، پرهزینه می‌باشد. بنابراین در عمل معمولاً این الگوریتم‌ها را چندین بار با حالات شروع مختلف بر روی نمونه‌ها اجرا می‌کنند و بهترین ترکیب به دست آمده، برای خروجی خوشه‌بندی استفاده می‌گردد (Jain, Murty and Flynn 1999). معروف‌ترین الگوریتم از نوع افزایشی، روش «میانگین کا»^۲ می‌باشد که مبنای بسیاری دیگر از الگوریتم‌های خوشه‌بندی است (Alpaydin 2004, 135).



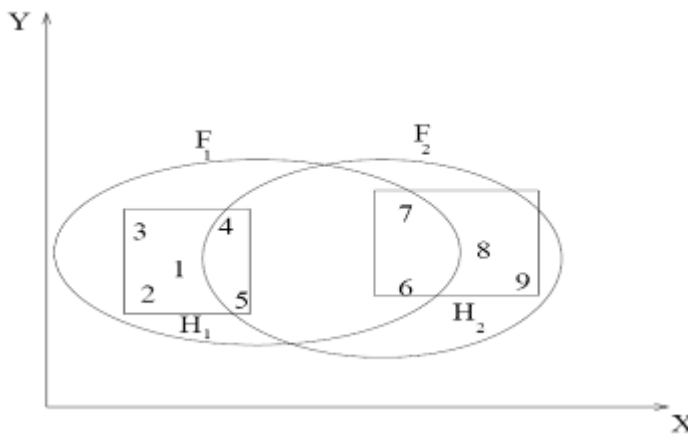
شکل ۲. نمونه‌ای از نتایج الگوریتم‌های افزایشی (Jain, Murty and Flynn 1999)

^۱ criterion function

^۲ K-means

۳-۲. خوشه‌بندی فازی

در راهبردهای خوشه‌بندی سنتی، هر نمونه به یک و تنها یک خوشه تعلق داشت. بنابراین خوشه‌های به‌دست آمده کاملاً از هم جدا و منفصل می‌باشند. در روش‌های خوشه‌بندی فازی، هر نمونه را می‌توان با یک تابع عضویت به هر یک از خوشه‌ها نسبت داد (Jain, Murty and Flynn 1999) و همانطور که در شکل ۳ نشان داده شده، یک عضو می‌تواند با توجه به توابع عضویت، همزمان در دو یا چند خوشه عضویت داشته باشد.



شکل ۳. نمونه‌ای از نتایج الگوریتم‌های فازی (Jain, Murty and Flynn 1999)

۳. ترسیم نقشه‌های علم

یکی از کاربردهای خوشه‌بندی، استفاده از این فنون در ترسیم نقشه‌های علم می‌باشد. ترسیم نقشه‌های علم از زیرشاخه‌های دانش علم‌سنجی است. در علم‌سنجی، از روش‌های آماری و اندازه‌گیری برای تعیین معیارهای رشد و توسعه علوم و سطوح گسترش آن‌ها و میزان تأثیر آن در جوامع مختلف بشری، استفاده می‌شود. بنا به تعریف «سن گویتا» (۱۹۹۲) علم‌سنجی علمی است که به ارزشیابی کمی آخرین پیشرفت هر مطلب علمی بنیادی و عوامل مؤثر در گسترش مستمر فعالیت‌های تحقیقاتی در آن زمینه خاص می‌پردازد. بر اساس نظر «ناگپاول»^۱ سه روند در مطالعات علم‌سنجی پیگیری شده است:

^۱ Nagpaul

۱. علم‌سنجی و سیاست‌های علمی و سیاست‌های فناورانه، شامل روش‌ها و حوزه‌های علم‌سنجی؛
۲. شناخت ساختار و پویایی‌های علم؛
۳. جنبه‌های منطقه‌ای و جغرافیایی توسعه علوم (Hood and Concepcins 2001).

بر اساس طبقه‌بندی فوق، یکی از روندهای علم‌سنجی، مطالعه ساختار علم و پویایی‌های آن می‌باشد. در این رویکرد، بعد از مطالعه رشته‌های مختلف، سرفصل‌هایی که تاکنون یا در بازه زمانی خاص، در آن رشته به آن‌ها پرداخته شده و حجم مطالعات در هر سرفصل و ارتباط بین این سرفصل‌ها، به صورت علمی و بر مبنای علوم ریاضی و آماری به روشنی تبیین می‌شود. در همین راستا نقشه‌های علمی در قالب شکل‌های گرافیکی به تفکیک حوزه‌های مختلف علوم و نمایش ارتباطات بین آن‌ها، به درک ساختارهای علوم کمک می‌کنند. نقشه‌ها به صورت دو بُعدی یا سه بُعدی، حوزه خاصی از علم را به تصویر می‌کشند. این نقشه‌ها در واقع در صدد پاسخگویی به سؤالات زیر می‌باشند (Leydesdorff 1987):

۱. ساختار طبیعی واحدهای تشکیل دهنده علوم چگونه است؟
 ۲. این واحدهای طبیعی چگونه به هم مرتبط می‌شوند؟
 ۳. چه نیروهایی ساختار واحدهای تشکیل دهنده علم و ارتباط بین آن‌ها را تعیین می‌کنند؟
 ۴. ساختارهای علمی چگونه در طول زمان در مقیاس خرد و کلان تغییر می‌کنند؟
- در نقشه‌های علم، حوزه‌های موضوعی که با هم ارتباط بیش‌تری دارند در فاصله نزدیک‌تر، و حوزه‌هایی که ارتباط کم‌تری دارند در فاصله بیش‌تر نمایش داده می‌شوند (Noyons 1999, 143). به اعتقاد برخی دانشمندان، تصویرسازی سه‌بعدی از رشته‌های مختلف علمی به ادراک سریع‌تر و بهتر در خصوص دانش مورد نظر کمک می‌کند (Small 2000). نقشه‌ها به روش‌های مختلف در سطوح مختلف، در حوزه‌های مکانی و زمانی مختلف و بر اساس مستندات مختلف ترسیم می‌شوند.
- فرایند اصلی ترسیم نقشه‌های علم شامل ۶ مرحله می‌باشد (Börner, Chen and Boyack 2003):

۱. استخراج داده^۱،
۲. تعریف واحدهای تحلیل^۲،
۳. انتخاب شاخص^۳،
۴. محاسبه شباهت‌های بین واحدها^۴،
۵. اولویت‌بندی^۵،
۶. استفاده از نتیجه برای تحلیل و تفسیر داده‌ها^۶.

جدول ۱. فرایند ترسیم نقشه‌های علم (Börner, Chen and Boyack 2003)

نمایش	ترسیم		انتخاب شاخص	تعریف واحدهای تحلیل	استخراج داده
	شباهت بین واحدها	دسته‌بندی			
تعامل:	کاهش ابعاد:	کمیت‌ها:	شمار / فراوانی:	انتخاب‌های	منابع جستجو:
مرور	روش‌های	استناد مستقیم	مشخصه‌ها (مثل واژگان)	رایج:	پایگاه‌های مختلف داده‌ای از قبیل:
درشت‌نمایی	ویژه‌مقدار / ویژه‌بردار	هم‌استنادی	استناد به نویسندگان	مجلات	ISI
فیلتر کردن	تحلیل عاملی و تحلیل	پیوندهای مختلط	هم‌استنادی	مدارک	Eng Index
پرس‌وجو	مؤلفه‌های اصلی	هم‌رخدادی واژگان	بر مبنای سال	نویسندگان	Medline
بررسی	مقیاس‌گذاری چند-بعدي	هم‌دستی	آستانه ^۷	واژه‌ها	پایگاه‌های داده‌ای
جزئیات	نقشه‌های مسیریاب	بُردارها:	شمار	مربوط به تحقیقات.	مربوط به ثبت اختراعات.
تحلیل	نقشه‌های خود سازمانده‌ی کننده	مدل بردار فضایی		پایگاه‌های داده‌ای	مربوط به ثبت اختراعات.
		تحلیل معنایی نهفته		مربوط به ثبت اختراعات.	و غیره
		همبستگی (در صورت نیاز):		شبهه مرزبندی:	توسط ارجاعات
	تحلیل خوشه‌ای	همبستگی پیرسون		توسط واژه‌ها	
	روش‌های کمی:	در مورد هریک از موارد فوق			
	مثلت‌بندی				
	جایگزاری نیرو-				
	هدایت شونده ^۸				

1 data extraction

2 definition of unit of analysis

3 selection of measures

4 calculation of a similarity between units

5 ordination

6 use of the resulting visualization

7 thresholds

8 force-directed placement (FDP)

اولین گام در هر فرایند ترسیم نقشه علمی، استخراج اطلاعات مناسب می‌باشد. راهبردهای مختلف جستجو در اینجا کاربرد دارند، اما مهم‌ترین نکته این که کیفیت نقشه‌هایی که ترسیم می‌شوند و حوزه‌هایی که به تصویر کشیده می‌شوند به صورت مستقیم وابسته به اطلاعاتی است که مبنای کار قرار می‌گیرند. تعداد مدارکی که برای ترسیم نقشه‌ها به کار می‌رود، می‌تواند از چند صد مدرک تا چندین هزار مدرک باشد (Börner, Chen and Boyack 2003).

انتخاب واحد تحلیل بستگی مستقیم به سؤالی دارد که در صدد پاسخگویی به آن هستیم؛ رایج‌ترین واحدها برای ترسیم نقشه‌ها عبارت‌اند از مجلات^۱، مدارک^۲ و واژه‌ها، و اصطلاحات توصیفگر^۳. هر یک از این واحدهای تحلیل، جنبه‌ای متفاوت از حوزه مورد مطالعه را نشان می‌دهند و انواع مختلف تحلیل را فراهم می‌آورند. مثلاً نقشه‌ای که برای تحلیل از مجلات استفاده می‌کند، تصویری کلان از علم مورد نظر را نمایش می‌دهد (Bassecoulard and Zitt 1999).

نقشه‌ها برای اهداف مختلفی مورد استفاده قرار می‌گیرند، از قبیل جستجوی مدارک، تحلیل حوزه‌ها، تصمیم‌گیری و سیاستگذاری علمی، ارزیابی داده‌های یافته‌شده، و همچنین مدیریت علم و فناوری و آگاهی از حوزه‌هایی که مورد رقابت قرار گرفته است (Boyack, Wylie and Davidson 2002).

شباهت‌های بین مدارک واحدها معمولاً با روش‌های مختلفی محاسبه می‌شود که رایج‌ترین آن‌ها عبارت‌اند از: ارتباطات ارجاعی^۴، شباهت‌های هم‌رخدادی^۵، و مدل بردار فضایی^۶.

روش‌های اولویت‌بندی متنوعی با توجه به کاربرد هر یک در ترسیم نقشه‌ها موجودند که مهم‌ترین آن‌ها عبارت‌اند از: تجزیه مقدار ویژه/ بردار ویژه^۷، تحلیل عاملی^۸،

¹ journals

² documents

³ descriptive terms or words Maps

⁴ citation linkages

⁵ co-occurrence similarities

⁶ vector space model

⁷ eigen value/eigenvector decomposition

⁸ factor analysis

مقیاس گذاری چندبُعدی^۱، تحلیل معنایی نهفته^۲، تحلیل خوشه‌ای^۳، و مثلث‌بندی^۴. طراحی تعاملی^۵ به تمام روش‌های نمایش و تصویرسازی اطلاعات اطلاق می‌شود که در جهت جستجو و پیمایش اثر بخش حوزه‌های گسترده اطلاعاتی می‌باشند. از جمله این روش‌ها می‌توان به انواع روش‌های فیلتر کردن^۶ اطلاعات، انواع روش‌های درشت‌نمایی^۷، و تغییر زاویه دید^۸ اشاره کرد (Börner, Chen and Boyack 2003). همان‌طور که اشاره شد، در گام سوم از فرایند ترسیم نقشه‌های علم، انواع الگوریتم‌های دسته‌بندی کاربرد دارند؛ یکی از مهم‌ترین روش‌های دسته‌بندی اطلاعات، روش تحلیل خوشه‌ها^۹ می‌باشد.

۴. کاربردهای خوشه‌بندی در ترسیم نقشه‌های علم

در بسیاری از تحقیقات بنیادین از خوشه‌بندی استفاده شده است. در سال ۱۹۷۳ «کارپنتر» و «نارین» به تشریح روش‌های مختلف خوشه‌بندی، از جمله روش‌های سلسله‌مراتبی و افزایی، برای خوشه‌بندی مجلات به‌منظور ترسیم نقشه‌های علم پرداختند (Carpenter and Narin 1973). در سال ۱۹۸۳، «گارفیلد» و «اسمال» در جریان تهیه نقشه علم برای علوم اجتماعی در این سال، از روش تک‌اتصال برای اجرای خوشه‌بندی اطلاعات استفاده کردند (Small and Garfield 1985). در سال ۱۹۸۷، «لیدسدرف» در مقاله‌ای مروی که به انواع روش‌های ترسیم نقشه علم پرداخته است در قسمت طبقه‌بندی و خوشه‌بندی اطلاعات، به فرایند خوشه‌بندی اطلاعات حاصل از جستجوی مقالات اشاره دارد (Leydesdorff 1987). «اسمال» و «سووینی» با انجام تغییر جزئی بر روی روش تک‌اتصال، از این روش برای ترسیم نقشه علمی که بر اساس مقالات طرح‌ریزی شده بود، استفاده کردند (Small and Sweeney 1985). در سال ۲۰۰۵، «بویاک»، «کلانوس» و

¹ multidimensional scaling

² latent semantic analysis

³ cluster analysis

⁴ triangulation

⁵ interaction design

⁶ filtering

⁷ zooming

⁸ distortion

⁹ cluster analysis

«بورنر» در تلاش برای ترسیم نقشه کلی علم از الگوریتم «میانگین‌کا» استفاده کرده‌اند (Boyack, Klavans and Borner 2005).

در سال ۲۰۰۶، «لیدسدورف» و «رافولز» با اشاره به تعداد ۶۱۶۴ مجله نمایه‌شده در مؤسسه آی‌اس‌آی در این تاریخ، روش خوشه‌بندی با الگوریتم تک‌اتصال را بهترین شیوه برای خوشه‌بندی این مجلات و مقالات موجود در این مقالات دانسته‌اند و در تحقیقات خود برای ترسیم نقشه یکپارچه علم بر اساس مستندات آی‌اس‌آی^۱، از این شیوه برای خوشه‌بندی بهره برده‌اند (Leydesdorff and Rafols 2006).

در ایران پژوهش‌های معدودی برای تدوین نقشه علمی حوزه‌های خاص انجام شده است که می‌توان به «ترسیم نقشه علمی نانو تکنولوژی در ایران» (محمدی ۱۳۸۷) و «ترسیم نقشه علم مدیریت شهری بر مبنای مستندات آی‌اس‌آی» (آقازاده ۱۳۸۸) اشاره کرد. در هر دوی این موارد از روش‌های خوشه‌بندی برای ترسیم نقشه‌های علم استفاده شده است. در ادامه، رایج‌ترین انواع الگوریتم‌های خوشه‌بندی که در فرایند ترسیم نقشه‌های علم کاربرد دارند به اختصار توضیح داده می‌شوند:

خوشه‌بندی با روش تک‌اتصال: این روش یکی از قدیمی‌ترین و ساده‌ترین روش‌های خوشه‌بندی است و جزو روش‌های خوشه‌بندی سلسله‌مراتبی محسوب می‌شود. به این روش خوشه‌بندی، شیوه «نزدیک‌ترین همسایه»^۲ نیز گفته می‌شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از معیار زیر استفاده می‌شود:

$$d_{AB} = \min_{i \in A, j \in B} d_{ij} \quad \text{فرمول ۱:}$$

که در آن، i یک نمونه داده متعلق به خوشه A، و j یک نمونه داده متعلق به خوشه B می‌باشد. در واقع در این روش شباهت بین دو خوشه، کم‌ترین فاصله بین یک عضو از یکی با یک عضو از دیگری است (Web 2002, 374).

خوشه‌بندی با روش اتصال کامل: این روش همانند روش تک‌اتصال جزو روش‌های خوشه‌بندی سلسله‌مراتبی و انحصاری محسوب می‌شود. به این روش خوشه‌بندی، شیوه

¹ Institute for Scientific Information (ISI)

² Nearest Neighbor

«دورترین همسایه»^۱ نیز گفته می‌شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از معیار زیر استفاده می‌شود:

$$d_{AB} = \max_{i \in A, j \in B} d_{ij} \quad \text{فرمول ۲:}$$

که در آن، i یک نمونه داده متعلق به خوشه A، و j یک نمونه داده متعلق به خوشه B می‌باشد. در واقع در این روش شباهت بین دو خوشه، بیشترین فاصله بین یک عضو از یکی با یک عضو از دیگری است (Web 2002, 367). گفتنی است که این نوع خوشه‌بندی در مواردی به کار می‌رود که فاصله به صورت استعاره‌ای از مفاهیمی مثل قدرت یا شدت رابطه استفاده شده و بیش تر شدن فاصله به معنای بالاتر رفتن شباهت باشد.

خوشه‌بندی با روش میانگین اتصال: این روش همانند تک‌اتصال، جزو روش‌های خوشه‌بندی سلسله‌مراتبی و انحصاری محسوب می‌شود. از آنجا که هر دو روش خوشه‌بندی تک‌اتصال و اتصال کامل بشدت به اختلال^۲ و اطلاعات اضافی و جانبی، حساس می‌باشند، این روش که محاسبات بیش تری دارد، پیشنهاد شد. در این روش برای محاسبه شباهت بین دو خوشه A و B از معیار زیر استفاده می‌شود:

$$d_{AB} = \frac{\sum_{i \in A, j \in B} d_{ij}}{N_A N_B} \quad \text{فرمول ۳:}$$

که در آن، i یک نمونه داده متعلق به خوشه A، و j یک نمونه داده متعلق به خوشه B می‌باشد. N_A تعداد اعضای خوشه A، و N_B تعداد اعضای خوشه B است. در واقع در این روش، شباهت بین دو خوشه، میانگین فاصله بین تمام اعضای یکی با تمام اعضای دیگری است (Web 2002, 374).

خوشه‌بندی به روش سلسله‌مراتبی تجمعی^۳: این الگوریتم در سه مرحله به خوشه‌بندی اطلاعات می‌پردازد، در مرحله اول، ماتریس مجاورتی می‌سازیم که فاصله هر جفت از نمونه‌ها را در آن نمایش می‌دهیم، و در واقع با هر نمونه به عنوان یک خوشه رفتار می‌کنیم.

¹ Furthest Neighbor

² noise

³ Hierarchical Agglomerative Clustering Algorithm

در مرحله دوم، با استفاده از ماتریس حاصل، جفتی را پیدا می‌کنیم که بسیار به هم شبیه هستند، کم‌ترین یا بیش‌ترین عدد در ماتریس؛ آن دو را در یک خوشه ادغام می‌کنیم و ماتریس مجاورت را به روز می‌کنیم تا مبین این عمل ادغام باشد. در مرحله سوم، شرط خاتمه‌بازینی می‌شود. شرط خاتمه این است که تمامی نمونه‌ها در یک خوشه باشند: در غیر این صورت به مرحله دو می‌رویم (Janin, Murty and Flynn 1999).

خوشه‌بندی با روش میانگین اتصال گروهی: این روش همانند روش تک‌اتصال جزو روش‌های خوشه‌بندی سلسله‌مراتبی و انحصاری محسوب می‌شود. به این روش، فاصله مرکزی^۱ نیز گفته می‌شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از معیار زیر استفاده می‌شود:

$$d_{AB} = d\left(\frac{\sum_{i \in A} X_i}{N_A}, \frac{\sum_{j \in B} X_j}{N_B}\right) \quad \text{فرمول ۴}$$

که در آن، X_i یک نمونه داده متعلق به خوشه A، X_j یک نمونه داده متعلق به خوشه B، N_A تعداد اعضای خوشه A و N_B تعداد اعضای خوشه B است. در واقع در این روش، شباهت بین دو خوشه، فاصله بین بُردار میانگین تمام اعضای یکی با بردار میانگین تمام اعضای دیگری است (Web 2002, 376).

خوشه‌بندی با روش نهبان: این روش نیز همانند روش تک‌اتصال جزو روش‌های خوشه‌بندی سلسله‌مراتبی و انحصاری محسوب می‌شود. این روش خوشه‌بندی برای کاهش تلفات ناشی از داده‌های دورافتاده^۲، از معیاری جدید برای محاسبه عدم شباهت بین خوشه‌ها استفاده می‌کند. در این روش، از مجموع مربعات تفاضل هر داده از یک خوشه با بردار میانگین آن خوشه، به‌عنوان معیاری برای سنجش یک خوشه استفاده می‌شود (He 1999).

روش خوشه‌بندی میانگین کا: این روش با وجود سادگی آن، یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. روش میانگین کا از انواع روش‌های افزایشی محسوب می‌شود (Alpaydin 2004, 137). برای این الگوریتم

^۱ centroid distance

^۲ outlier

شکل‌های مختلفی بیان شده، ولی همه آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

◇ به دست آوردن نقاطی به عنوان مراکز خوشه‌ها، که این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.

◇ نسبت دادن هر نمونه داده به یک خوشه، که آن داده کم‌ترین فاصله تا مرکز آن خوشه را دارا باشد.

در نوع ساده‌ای از این روش، ابتدا به تعداد خوشه‌های مورد نیاز، نقاطی به صورت تصادفی انتخاب می‌شود. سپس داده‌ها با توجه به میزان نزدیکی شباهت، به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها، مراکز جدیدی برای آن‌ها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود.

الگوریتم خوشه‌بندی ال‌بی‌جی^۱: در این روش، الگوریتم ابتدا تمام داده‌ها را به صورت یک خوشه در نظر می‌گیرد و با اجرای الگوریتم برای تعداد یک خوشه، یک بردار مرکز محاسبه می‌کند. سپس این بردار را به دو بردار می‌شکند و داده‌ها را با توجه به این دو بردار خوشه‌بندی می‌کند. در مرحله بعد این دو نقطه به چهار نقطه شکسته می‌شوند و الگوریتم ادامه پیدا می‌کند تا تعداد خوشه مورد نظر تولید شوند.

الگوریتم‌های مبتنی بر نظریه گراف: این نوع الگوریتم‌ها از نوع الگوریتم‌های تقسیمی هستند. شناخته‌ترین الگوریتم خوشه‌بندی تقسیمی بر پایه نظریه گراف، ساختن درخت کمینه^۲ از نمونه‌ها می‌باشد (Zahn 1971). در این الگوریتم سعی می‌شود که یال‌های با بزرگ‌ترین طول‌ها حذف شوند تا خوشه‌های دلخواه ایجاد گردند.

۵. مورد کاوی: استفاده از خوشه‌بندی در ترسیم نقشه علم مدیریت شهری

هدف از ترسیم نقشه علم مدیریت شهری، پاسخ به این پرسش بوده است که «دانش

^۱ LBG (Linde-Buzo-Gray)

^۲ minimal

مدیریت شهری از چه زیرحوزه‌های موضوعی تشکیل شده، و ارتباط این زیرحوزه‌ها با یکدیگر چگونه می‌باشد؟» به این منظور پایگاه اطلاعاتی «آی‌اس‌آی» به‌عنوان منبع استخراج اطلاعات انتخاب شد. آی‌اس‌آی، پایگاه داده‌ای برای فهرست نمودن و پوشش دادن مهم‌ترین مجلات علمی منتشره در دنیا به منظور تبادل اطلاعات میان پژوهشگران مختلف می‌باشد، که در حال حاضر مهم‌ترین و جامع‌ترین پایگاه اطلاعاتی استنادی در سراسر جهان است (علیجانی و کرمی ۱۳۸۷، ۱۳۳).

برای استخراج نقشه علم مدیریت شهری از آی‌اس‌آی، در پنج گام اقدام شد: (۱) طراحی زنجیره جستجو؛^۱ (۲) انجام جستجو بر اساس زنجیره جستجوی طراحی شده؛ (۳) تجزیه و تحلیل نتایج جستجو؛ (۴) طراحی ماتریس شدت ارتباط زیرحوزه‌های موضوعی؛ (۵) استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی و ماتریس شدت ارتباطات زیرحوزه‌های موضوعی برای ترسیم نقشه علم مدیریت شهری.

پنج گام فوق به‌صورت عملیاتی برای ترسیم نقشه علم مدیریت شهری، شامل مراحل زیر می‌باشد (آقازاده ۱۳۸۸):

۱. طراحی راهبرد جستجو: کلیدواژه‌های مورد نیاز برای انجام جستجو از «دانشنامه مدیریت شهری و روستایی» که با همکاری «وزارت علوم، تحقیقات و فناوری» و «سازمان شهرداری‌ها و دهیاری‌های کشور» (وابسته به وزارت کشور) تهیه شده است، استخراج شد. با مراجعه به ویرایش اول «دانشنامه مدیریت شهری و روستایی»، مشاهده می‌شود که این دانشنامه با حجم بیش از ۹۰۰ صفحه مطلب درباره ۳۵۰ مدخل مرتبط به مباحث مختلف تخصصی مدیریت شهری و روستایی، تنظیم شده است.
۲. انجام جستجو: با استفاده از کلیدواژه‌های استخراج شده از «دانشنامه مدیریت شهری و روستایی»، عملیات جستجو در وبگاه آی‌اس‌آی صورت پذیرفت. محدوده زمانی برای جستجوی مقالات، بین سال‌های ۲۰۰۰ تا ۲۰۰۸ میلادی در نظر گرفته شد.
۳. تجزیه و تحلیل نتایج: بعد از انجام عملیات جستجو و تجزیه و تحلیل نتایج با استفاده از نرم‌افزار تحلیلگر سایت آی‌اس‌آی و نرم‌افزار آماری اسپ‌اس‌اس، ۳۶ طبقه‌بندی موضوعی به‌عنوان زیرحوزه‌های موضوعی رشته مدیریت شهری تعیین شد.

¹ Search String

۴. طراحی ماتریس تجزیه و تحلیل: بر اساس اطلاعات مربوط به میزان مقالات مشترک در زیرحوزه‌های موضوعی و نرمال‌سازی بر مبنای درصد مقالات مشترک در زیرحوزه‌های موضوعی، «ماتریس نرمال ارتباطات حوزه‌های موضوعی مختلف» به دست آمد.

۵. ترسیم نقشه: برای طبقه‌بندی ۳۶ زیرحوزه موضوعی و به دست آوردن نواحی نقشه، از الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی استفاده گردید. برای ترسیم نقشه و مصورسازی، نرم‌افزار «پاژک» به کار گرفته شد. اندازه گره‌ها در ترسیم نقشه بر مبنای حجم مقاله در هر زیرحوزه موضوعی محاسبه شد و وزن یال‌ها (خطوط بین گره‌ها) معرف درصد مقالات مشترک در زیرحوزه‌های موضوعی مرتبط در نظر گرفته شد.

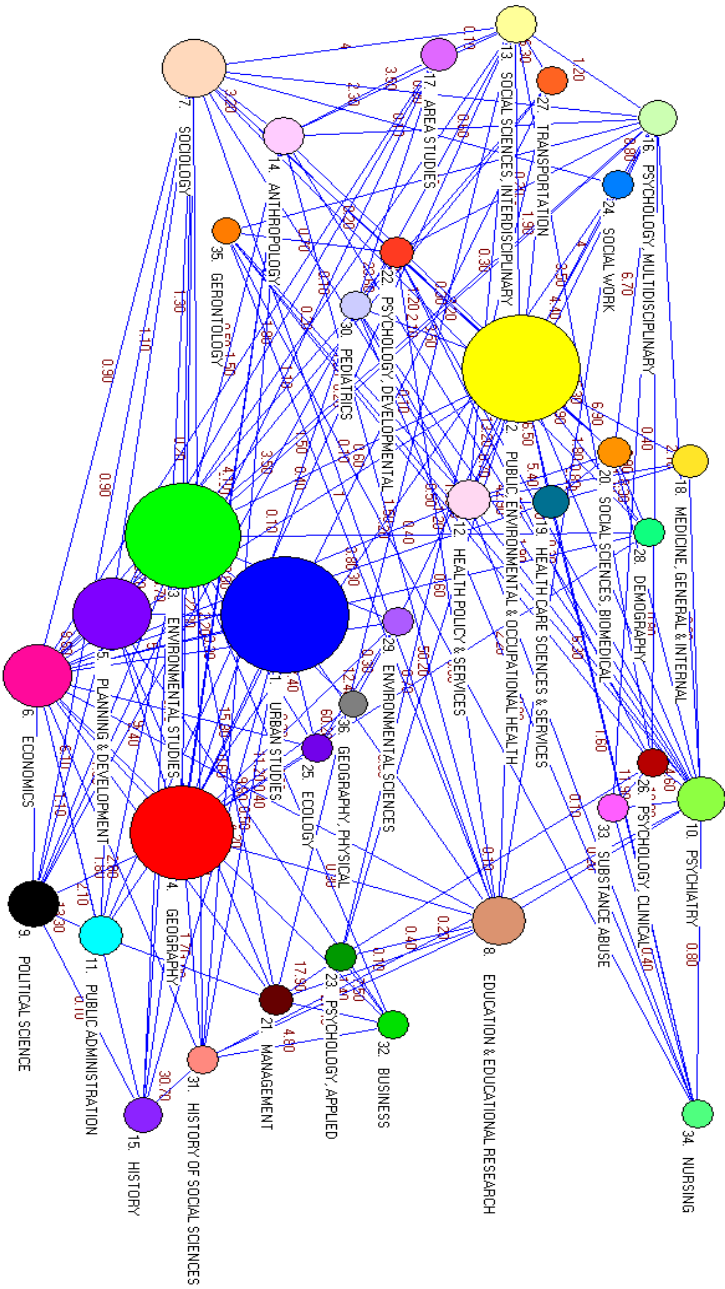
یافته‌های پژوهشی نشان می‌دهد که نقشه علم مدیریت شهری شامل ۳۶ زیرحوزه موضوعی و ۱۸ ناحیه می‌باشد. ۴۴ ناحیه دیگر نیز که قابلیت پیوستن به مجموعه زیرحوزه‌های موضوعی علم مدیریت شهری را دارند، مشخص شد. نقشه علم مدیریت شهری در دو سطح ترسیم شد: سطح اول که متشکل از ۳۶ زیرحوزه موضوعی و نمایش میزان ارتباطات این زیرحوزه‌ها می‌باشد، و سطح دوم که با استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی^۱، ۳۶ زیرحوزه موضوعی در ۷ خوشه نواحی نقشه و ۱۱ زیرحوزه موضوعی مستقل، سازماندهی شد. در سطح اول نقشه، اطلاعات بسیار زیاد و مناسبی در قالب یک نقشه ارائه شده است. در این سطح با توجه به این که اطلاعات بسیار جامع و مفید می‌باشد، اطلاعات ارائه شده تا حدی پیچیده و برای افراد غیرمتخصص، تا حدی نامفهوم می‌باشد. از این رو برای ساده‌سازی بیش‌تر، اطلاعات در طبقه‌بندی‌های کلی‌تر ارائه شد. از آنجا که طبقه خاصی از پیش تعیین نشده بود، برای ساده‌سازی بیش‌تر و ادراک بهتر، ۳۶ زیرحوزه و میزان ارتباطات آن‌ها تحت الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی، مورد تجزیه و تحلیل قرار گرفت (جدول ۱).

بر اساس الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی که پیش‌تر توضیح داده شد، فرایند خوشه‌بندی با ساختن ماتریس مجاورتی آغاز شد (جدول ۳ که فاصله هر دو نمونه در آن نمایش داده شده).

^۱ Hierarchical Agglomerative Clustering Algorithm

جدول ۱. فهرست زیرحوزه‌های موضوعی در نقشه علم مدیریت شهری

حوزه موضوعی	اختصار	فراوانی	درصد	درصد تجمعی
URBAN STUDIES	US	۵۶۴۲	۸/۶	۸/۶
PUBLIC, ENVIRONMENTAL AND OCCUPATIONAL HEALTH ENVIRONMENTAL STUDIES	PEOH	۵۰۱۷	۷/۷	۱۶/۳
GEOGRAPHY	ES	۴۹۰۸	۷/۵	۲۳/۹
PLANNING and DEVELOPMENT	G	۴۲۴۲	۶/۵	۳۰/۴
ECONOMICS	PD	۲۹۶۷	۴/۶	۳۴/۹
SOCIOLOGY	E	۲۵۲۲	۳/۹	۳۸/۸
EDUCATION and EDUCATIONAL RESEARCH	S	۲۲۰۴	۳/۴	۴۲/۲
POLITICAL SCIENCE	EER	۱۷۷۵	۲/۷	۴۴/۹
PSYCHIATRY	PS	۱۶۴۹	۲/۵	۴۷/۴
PUBLIC ADMINISTRATION	P	۱۵۰۷	۲/۳	۴۹/۸
HEALTH POLICY and SERVICES	PA	۱۲۳۱	۱/۹	۵۱/۶
SOCIAL SCIENCES, INTERDISCIPLINARY	HPS	۱۲۱۱	۱/۹	۵۳/۵
ANTHROPOLOGY	SSI	۱۱۷۶	۱/۸	۵۵/۳
HISTORY	A	۱۰۶۹	۱/۶	۵۶/۹
PSYCHOLOGY, MULTIDISCIPLINARY	H	۱۰۳۵	۱/۶	۵۸/۵
AREA STUDIES	PM	۱۰۰۶	۱/۵	۶۰/۱
MEDICINE, GENERAL and INTERNAL	AS	۹۲۹	۱/۴	۶۱/۵
HEALTH CARE SCIENCES and SERVICES	MGI	۸۶۷	۱/۳	۶۲/۸
SOCIAL SCIENCES, BIOMEDICAL	HSS	۸۲۷	۱/۳	۶۴/۱
MANAGEMENT	SSB	۷۶۷	۱/۲	۶۵/۳
PSYCHOLOGY, DEVELOPMENTAL	M	۷۳۴	۱/۱	۶۶/۴
PSYCHOLOGY, APPLIED	PSD	۷۱۴	۱/۱	۶۷/۵
SOCIAL WORK	PSA	۶۸۰	۱	۶۸/۵
ECOLOGY	SW	۶۷۵	۱	۶۹/۶
PSYCHOLOGY, CLINICAL	ECO	۶۷۲	۱	۷۰/۶
TRANSPORTATION	PC	۶۶۴	۱	۷۱/۶
DEMOGRAPHY	TR	۶۶۳	۱	۷۲/۶
ENVIRONMENTAL SCIENCES	DE	۶۳۵	۱	۷۳/۶
PEDIATRICS	ESC	۶۱۵	۰/۹	۷۴/۶
HISTORY OF SOCIAL SCIENCES	PED	۶۱۲	۰/۹	۷۵/۵
BUSINESS	HOSS	۶۰۶	۰/۹	۷۶/۴
SUBSTANCE ABUSE	BU	۵۸۶	۰/۹	۷۷/۳
NURSING	SA	۵۶۶	۰/۹	۷۸/۲
GERONTOLOGY	NU	۵۶۵	۰/۹	۷۹/۱
GEOGRAPHY, PHYSICAL	GRO	۵۴۷	۰/۸	۷۹/۹
	GP	۵۳۶	۰/۸	۸۰/۷



شکل ۴. نقشه علم مدیریت شهری قبل از اتمام خوشه‌بندی

منظور از فاصله، درصد مقالات مشترک بین دو زیرحوزه موضوعی می‌باشد. اعداد بزرگ‌تر در جدول به معنای درصد اشتراک مقالات بیش‌تر و فاصله کم‌تر می‌باشد. در واقع با هر نمونه به‌عنوان یک خوشه رفتار شد. سپس با استفاده از ماتریس حاصل، جفت‌هایی که بسیار به هم شبیه بودند، پیدا شدند. جفت‌های مشابه در یک خوشه ادغام، و بعد از هر ادغام، ماتریس مجاورت به روز می‌شد تا مبین این ادغام باشد. شرط خاتمه حضور تمامی نمونه‌ها در یک خوشه تعیین شد. در غیر این صورت، بازگشت به مرحله دو انجام می‌شد و این کار تا وقتی که تمام گره‌ها به یک خوشه واحد پیوندند، ادامه می‌یافت. معیار شناسایی حوزه‌های کلی‌تر که آن‌ها را نواحی نقشه می‌نامیم، استقلال خوشه‌های تشکیل‌دهنده این حوزه‌ها بود، به این نحو که خوشه تشکیل‌دهنده هر یک از این حوزه‌ها، هر قدر زودتر ایجاد شود، و هر قدر دیرتر ادغام شود، نشان‌دهنده استقلال بیش‌تر در این حوزه‌ها می‌باشد. هر چه خوشه‌ای زودتر تشکیل شود، نشان‌دهنده همبستگی بیش‌تر هسته تشکیل‌دهنده خوشه می‌باشد و هر قدر دیرتر ادغام شود، نشان‌دهنده همبستگی بیش‌تر اعضا می‌باشد. معیار شناسایی برای ترسیم خط مبنا، ادغام دو خوشه شماره ۱ و شماره ۲ بود. تا پیش از این ادغام، خوشه‌ها به‌صورت مستقل عضوگیری، و زیرحوزه‌های موضوعی مرتبط را جذب می‌کردند؛ اما با این ادغام، دو حوزه مستقل با هم ادغام شدند، که از تجانس کافی برخوردار نبودند. نتیجه حاصل از اجرای الگوریتم فوق به‌صورت خلاصه در نمودار دندان‌های شکل حاصل از اجرای الگوریتم به تصویر کشیده شده (شکل ۵). در این شکل، نحوه ادغام خوشه‌ها و فاصله‌های ادغام به تفصیل مشخص شده است.

۶. یافته‌های پژوهش

بر اساس روش فوق، ۳۶ زیرحوزه موضوعی تشکیل‌دهنده سطح ۱ نقشه علم مدیریت شهری به ۱۸ دسته کلی‌تر تقسیم می‌گردند که ۷ حوزه، نواحی نقشه را تشکیل می‌دهند و ۱۱ حوزه، از حوزه‌های مستقل می‌باشند. در نمودار دندان‌های شکل حاصل، این ۷ ناحیه با عناوین خوشه ۱ تا ۷ مشخص شده‌اند.

نام ۱۸ ناحیه نقشه، به ترتیب فراوانی مقالات در جدول ۲ آمده است.

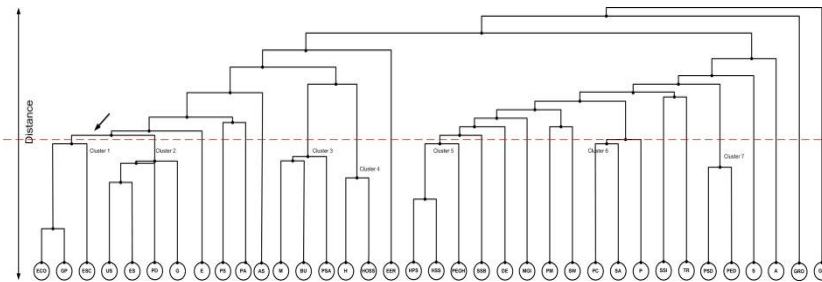
جدول ۲. نواحی نقشه علم مدیریت شهری بعد از اعمال الگوریتم خوشه‌بندی

ردیف	نام ناحیه / زیرحوزه موضوعی	درصد مقالات
۱	مجموعه دانش شهر و آبادانی شهری	۳۳/۷۳
۲	مجموعه دانش پزشکی و بهداشت شهری	۱۷/۷۳
۳	حوزه موضوعی اقتصاد	۶/۵۰
۴	جامعه‌شناسی	۵/۷۰
۵	مجموعه دانش مدیریت دولتی و علوم سیاسی	۵/۴۸
۶	مجموعه دانش روانشناسی	۵/۳۰
۷	حوزه موضوعی تحقیقات آموزش و آموزشی	۴/۶۰
۸	مجموعه دانش مدیریت	۳/۸۰
۹	مجموعه دانش مربوط به بوم‌شناسی شهری	۳/۴۷
۱۰	مجموعه دانش تاریخ	۳/۱۲
۱۱	حوزه موضوعی مردم‌شناسی	۲/۷۰
۱۲	حوزه موضوعی روانشناسی بین‌رشته‌ای	۲/۷۰
۱۳	مجموعه دانش پزشکی و روانپزشکی کودکان	۲/۵۲
۱۴	حوزه موضوعی مطالعات ناحیه‌ای	۲/۴۰
۱۵	حوزه موضوعی امور اجتماعی	۱/۷۰
۱۶	حوزه موضوعی ترابری یا حمل‌ونقل	۱/۷۰
۱۷	پرستاری	۱/۴۰
۱۸	پیری‌شناسی	۱/۴۰

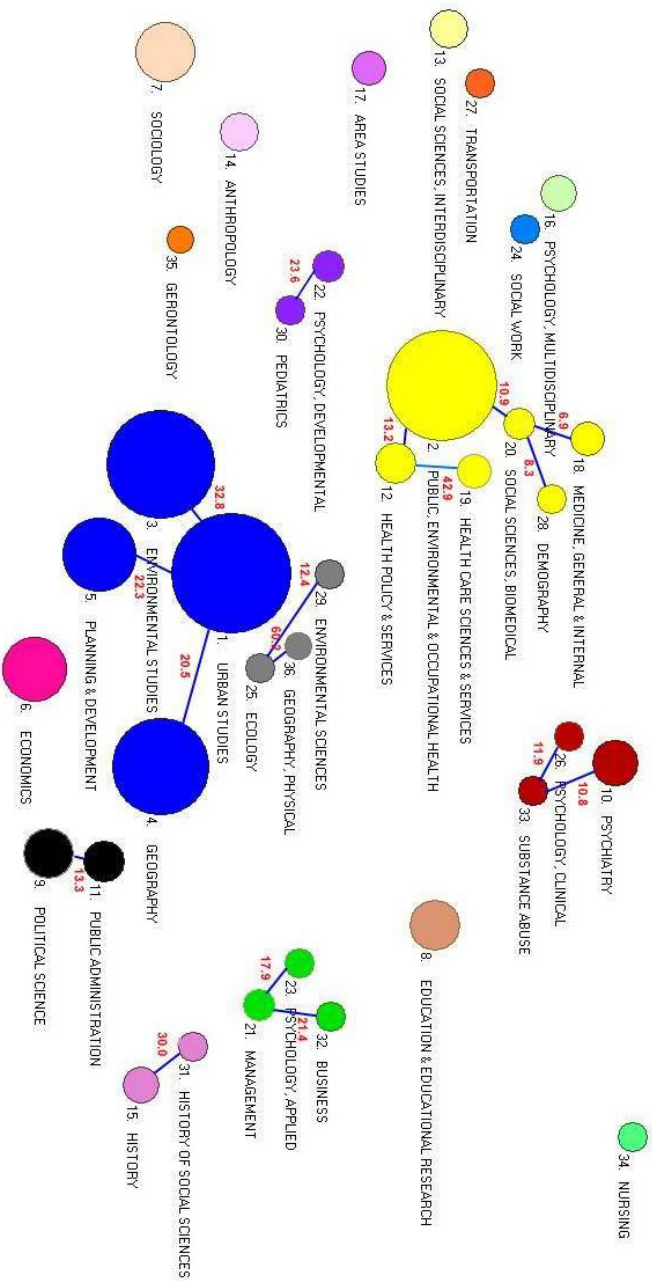
در جدول ۳، ماتریس ارتباطات بین زیرحوزه‌های مختلف نقشه علم نمایش داده شده است. اعداد ماتریس بیانگر درصد مقالات مشترک بین دو زیرحوزه موضوعی می‌باشند. این ماتریس ورودی الگوریتم خوشه‌بندی می‌باشد.

جدول ۳. ماتریس نرمال ارتباطات حوزه‌های موضوعی مختلف بر مبنای تعداد مقالات مشترک

	US	PEOH	ES	G	PD	E	S	EER	PS	P	PA	HPS	SSI	A	H	PM	AS	MGI	HSS	SSB	M	PSD	PSA	SW	ECO	PC	TR	DE	ESC	PED	HOSS	BU	SA	NU	GRO	GP	
US	08	00	328	158	223	50	00	35	00	00	00	00	00	00	39	00	00	00	00	00	00	00	00	00	00	74	00	00	00	00	42	00	00	00	00	00	
PEOH	00	00	01	00	00	01	02	22	00	24	00	132	21	00	00	40	00	69	65	109	00	32	02	35	00	00	19	18	13	35	00	00	00	16	00	00	
ES	328	01	00	05	107	57	13	00	06	00	24	00	00	05	03	21	00	00	00	00	06	00	00	00	106	00	17	04	54	00	00	02	00	00	00	00	87
G	158	00	205	00	00	39	02	03	18	00	00	00	00	03	00	14	00	00	00	00	00	00	00	98	00	06	07	02	00	17	00	00	00	00	00	112	
PD	221	00	107	90	00	00	88	11	00	61	00	78	00	02	00	00	18	00	00	00	02	00	00	00	00	00	00	00	00	00	00	05	00	00	00		
E	50	01	57	39	88	00	19	04	23	00	11	07	00	00	21	02	15	00	07	00	03	00	00	31	00	11	02	33	00	23	18	00	00	00	00		
S	00	02	13	02	11	09	01	00	09	00	01	40	32	07	02	00	00	07	00	00	00	01	00	01	00	00	12	00	00	08	00	00	00	00	00	00	
EER	35	22	00	03	00	04	10	00	00	02	00	00	00	00	12	00	00	00	00	00	00	01	06	04	03	00	00	00	00	14	00	00	00	00	03	00	
PS	00	00	06	18	61	23	09	00	00	133	00	35	00	01	00	41	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
P	00	24	00	00	00	00	00	00	00	00	63	00	07	00	21	00	02	00	08	00	23	00	00	00	46	00	00	19	02	00	108	08	30	00	00	00	
PA	00	00	24	00	78	11	00	00	133	00	00	00	00	00	00	00	00	00	00	18	00	00	38	00	00	00	00	00	00	00	00	00	00	00	00	00	00
HPS	00	132	00	00	00	07	01	00	00	63	00	00	03	00	00	44	00	05	429	54	00	00	00	00	00	00	00	01	00	00	00	01	12	00	00	01	12
SSI	00	03	00	00	02	00	40	00	35	00	04	03	00	00	15	00	12	01	00	00	15	00	00	00	53	63	00	00	00	00	00	00	00	00	00	00	00
A	00	00	05	03	00	00	32	12	00	07	00	00	35	00	22	06	23	00	00	00	00	00	00	00	00	00	00	00	00	00	15	00	00	00	00	00	00
H	39	00	03	00	00	00	21	07	00	01	00	00	00	00	22	00	00	00	00	00	00	00	00	00	00	00	00	00	30	07	00	00	00	00	00	00	00
PM	00	40	21	00	00	02	02	00	00	21	00	44	12	06	00	00	00	00	67	00	02	00	88	00	04	00	00	00	00	00	00	09	00	11	00	00	00
AS	00	00	21	00	00	14	18	15	00	00	41	00	00	00	01	23	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
MGI	00	69	00	00	00	00	00	00	02	00	05	00	00	00	00	00	00	00	17	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
HSS	00	65	00	00	00	07	00	00	00	00	00	429	00	00	00	00	00	00	17	00	08	00	00	00	00	00	00	00	00	00	00	00	00	00	00	15	00
SSB	00	109	00	00	00	00	00	00	08	00	54	00	30	00	67	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
M	00	00	06	00	02	03	00	01	00	00	18	00	15	00	00	00	00	00	00	00	00	00	179	00	00	00	00	00	00	00	00	214	00	00	00	00	00
PSD	00	32	00	00	00	00	06	00	23	00	00	00	00	00	02	00	00	00	00	00	00	00	00	00	00	00	00	00	00	236	00	00	00	00	00	07	00
PSA	00	02	00	00	00	00	04	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	75	00	00	00	00	00	00	
SW	00	35	00	00	00	00	01	03	00	00	38	00	00	00	00	98	00	00	00	00	169	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
ECO	74	00	106	98	00	31	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	124	00	00	00	00	00	00	00	602	
PC	00	00	00	00	00	00	00	00	46	00	00	53	00	00	04	00	00	00	00	00	00	00	41	01	00	00	00	00	00	00	119	00	05	00	00	00	
TR	00	19	17	06	00	11	00	00	00	00	00	63	00	00	00	00	00	00	00	00	00	00	12	00	00	00	00	00	00	00	00	00	00	00	00	00	
DE	00	18	04	07	00	02	12	00	00	00	00	00	00	00	00	00	00	00	83	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
ESC	00	13	54	02	00	33	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	124	00	00	00	00	00	00	00	00	00	00	00	00	00	03
PED	00	35	00	00	00	00	00	00	00	19	00	01	00	00	00	00	00	00	00	00	00	236	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
HOSS	42	00	00	17	00	28	08	14	00	02	00	00	00	15	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	48	00	00	00	00	00	00	
BU	00	00	02	00	05	18	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	75	00	00	00	00	00	00	00	48	00	00	00	00	00	00	
SA	00	00	00	00	00	00	00	00	109	00	00	00	00	00	00	00	00	00	00	00	00	212	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
NU	00	16	00	00	00	00	00	00	00	08	00	01	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	06
GRO	00	00	00	00	00	00	00	03	00	00	30	00	12	00	00	11	00	15	00	00	00	07	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
GP	00	00	87	112	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	602	00	00	00	03	00	00	00	00	00	00	00	00



شکل ۵. نمودار دندانه‌ای شکل مربوط به ترسیم نقشه علم مدیریت شهری



شکل ۱. خوشه‌های شناسایی شده بر اساس روش بزرگ‌ترین

۷. بحث و نتیجه‌گیری

در مقاله حاضر، به بررسی روش‌های مختلف خوشه‌بندی اطلاعات پرداخته شد. انتخاب روش خوشه‌بندی به مسئله پژوهش، ماهیت داده‌هایی که تحت فرایند خوشه‌بندی قرار می‌گیرند، و معنای داده‌ها وابسته می‌باشد. در ترسیم نقشه علم مدیریت شهری، از آنجا که برای خوشه‌بندی و دسته‌بندی مجموعه ۳۶ عضوی نقشه علم مدیریت شهری، معیاری در دسترس نبود، تصمیم بر آن شد که از یک الگوریتم سلسله‌مراتبی استفاده شود تا تمام مراحل تشکیل خوشه‌ها را به تفصیل داشته باشیم و سپس بر اساس نمودار دندانه‌ای شکل طراحی شده، سطح برش در جهت ایجاد خوشه‌ها مشخص شود؛ بنابراین الگوریتم خوشه‌بندی سلسله‌مراتبی تجمعی کامل اجرا شد و سپس بر مبنای ادغام دو خوشه بزرگ و نامتجانس، خوشه‌ها شناسایی شد.

برای بررسی پایایی نتایج پژوهش پیشنهاد می‌شود که داده‌های حاصل با الگوریتم‌های خوشه‌بندی مختلف مورد تجزیه و تحلیل قرار بگیرد. این امر همچنین منجر به شناسایی معیاری برای مقایسه انواع الگوریتم‌های خوشه‌بندی می‌شود.

۸. فهرست منابع

- آقازاده، فتاح. ۱۳۸۸. ترسیم نقشه علم مدیریت شهری بر مبنای مستندات آی‌اس‌آی. پایان‌نامه کارشناسی ارشد، دانشگاه تهران.
- حاج‌احمدی، امیرحسین. ۱۳۸۵. مبانی خوشه‌بندی. دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه امیرکبیر. <http://ceit.aut.ac.ir/~shiry/lecture/machine-learning/tutorial/clustering/Introduction.htm> (تاریخ دسترسی ۱۳۸۷/۴/۵)
- سن‌گوپتا، آی.ان. ۱۹۹۲. مروری بر کتاب‌سنجی، اطلاع‌سنجی، علم‌سنجی و کتابخانه‌سنجی. ترجمه مهرداد وزیرپور کشمیری. ۱۳۷۲. علوم اطلاع‌رسانی ۱۰(۲): ۳۸-۵۸.
- علیجانی، رحیم، و نورالله کرمی. ۱۳۸۷. مطالعات سنجش کمی. تهران: چاپار.
- محمدی، احسان. ۱۳۸۷. ترسیم نقشه علمی نانو فناوری در ایران. پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران.

- Alpaydin, E. 2004. *Introduction to Machine Learning*. Cambridge, MA: The MIT Press.
- Anderberg, M. R. 1973. *Cluster Analysis for Applications*. New York: Academic Press, Inc.
- Backer, E. 1995. *Computer-Assisted Reasoning Cluster Analysis*. Hertfordshire: Prentice Hall International.

- Bassecoulard, E. and M. Zitt. 1999. Indicators in a research institute: A multi-level classification of journals. *Scientometrics* 44(3): 323-345.
- Börner, K., C. Chen and K. W. Boyack. 2003. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology* 37: 179-225.
- Boyack, K.W., R. Klavans and K. Borner. 2005. Mapping the backbone of science. *Scientometrics* 64(3): 351-374.
- Boyack, K., B.N. Wylie and G. Davidson. 2002. Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology* 53(9):763-774.
- Carpenter, P. M. and F. Narin. 1973. Clustering of Scientific Journals. *Journal of the American Society for Information Science* 24(6): 425- 436.
- Duda, R., P. Hart and D. Stork. 2000. *Pattern Classification And Scene Analysis*. Hoboken: John Wiley and Sons.
- Edelstein, H.A. 1999. *Introduction to Data Mining and Knowledge Discovery*. Maryland: Two Crows Corporation.
- Garfield, E. 1963. Citation indexes in sociological and historical research. *American Documentation* 14(4): 289-291.
- Hartigan, J. 1975. *Clustering Algorithms*. New York: John Wiley and Sons.
- He, Q. 1999. *A Review of Clustering Algorithms as Applied in IR*. Masters thesis. Graduate School of Library and Information Science, University of Illinois.
- Hjorland, B. 1997. *Information Seeking and Subject Representation: An Activity-Theoretical*. Santa Barbara: Green wood press.
- Hood, W. and W. Concepcins. 2001. The Literature of Bibliometrics, Scientometrics and Informetrics. *Scientometrics*: 291-314.
- Jain, A.K. and R.C. Dubes. 1998. *Algorithms for Clustering Data*. New Jersey: Prentice-Hall.
- Jain, A. and P. Flynn. 1996. *Image segmentation using clustering*. Hoboken: IEEE Press.
- Jain, A., M. Murty and P. Flynn. 1999. Data Clustering: A Review. *ACM Computing Surveys* 31(3): 264 - 323.
- Mahamed, G.H. Omran, Andries P. Engelbrecht and Ayed Salman. 2007. An overview of clustering methods. *Intelligent Data Analysis* 11(6): 583-605.
- Leydesdorff, L. 1987. Various Methods for Mapping of Science. *Scientometric* 11(5-6): 295-324.
- Leydesdorff, L. and I. Rafols. 2006. A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology* 60: 348-362.
- Noyons, E.C. 1999. *Bibliometric Mapping as a Science Policy and Research Management Tool*. Leiden: DSWO Press.
- Otlet, P. 1918. *Transformations in the Bibliographical Apparatus of the Sciences. The International Organization and Dissemination of Knowledge: Selected Essays of Paul Otlet*. Elsevier.
- Price, D.D. 1965. Networks of scientific papers. *Science* 149: 510-519.
- Rasmussen, E. 1992. *Clustering algorithms in In Information Retrieval: Data Structures and Algorithms*. Hertfordshire: Prentice-Hall.
- Small, H. 2000. passage through science: Crossing disciplinary boundaries. *Library Trends* 48(1): 72-108.

- Small, H. and E. Garfield. 1985. The geography of science: disciplinary and national mappings. *Journal of Information Science* 11: 147-159.
- Small, H. and E. Sweeney. 1985. Clustering the Science Citation Index Using Co-Citations I. A Comparison of Methods. *Scientometrics* 7: 392-409.
- Spath, H. 1980. *Cluster Analysis Algorithms for Data Reduction and Classification*. New York: Halsted Press.
- Web, A. 2002. *Statistical Pattern Recognition*. Hoboken: John Wiley and Sons.
- Zahn, C. 1971. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers* 20(1): 68-86.

Applying Clustering Methods in Drawing Maps of Science: Case Study of the Map For Urban Management Science

Muhammad Aboui Ardakan

PhD in Management, Asst. Prof.,
University of Tehran

Hassan Abedi Jafari

PhD in Management, Asst. Prof.,
University of Tehran

Fatah Aghazadeh DehDeh*

MBA, Urban Management, University of Tehran

Information
Sciences
& Technology

Abstract: The present paper offers a basic introduction to data clustering and demonstrates the application of clustering methods in drawing maps of science. All approaches towards classification and clustering of information are briefly discussed. Their application to the process of visualization of conceptual information and drawing of science maps are illustrated by reviewing similar researches in this field. By implementing aggregated hierarchical clustering algorithm, which is an algorithm based on complete-link method, the map for urban management science as an emerging, interdisciplinary scientific field is analyzed and reviewed.

Keywords: information clustering; scientometry; science maps; urban management science

Iranian Research Institute
for Science and Technology

(IRANDOC)

ISSN 1735-5206

eISSN 2008-5583

Indexed in LISA & SCOPUS

Vol. 25 | No. 3 | pp: 347-371

Spring 2010

* Corresponding Author fatah.aghazadeh@gmail.com