



## Clustering of Bursts in Water Distribution Networks for Prioritization of the Rehabilitation Area

A.M. Aryayee<sup>1</sup>, M. Jalili Ghazizadeh<sup>2\*</sup>,  
and I. Moslehi<sup>3</sup>

### Abstract

Water supply networks are directly related to the economic and social development of urban communities. Due to the rapid increase in population and urbanization, these networks should be properly maintained and rehabilitated. In this paper, an unsupervised learning method with OPTICS density-based clustering algorithm is used to determine the priority areas for rehabilitation. 361 burst event data in one of the Mashhad water network zones were analyzed. The algorithm identified a total number of 16 clusters at 5 different levels of importance. These 16 clusters, which are considered sub-clusters for higher levels, were compared in terms of burst rate. Three clusters with failure rates of 79.1, 62.2, and 50.1 (failures/100 km/year) were introduced as the main priorities for rehabilitation and renovation, respectively. The average burst rate in the whole network was 14.8. The identified clusters were examined in terms of pipe material and diameter and their most affecting factors on the burst rate. The obtained results showed that the OPTICS clustering algorithm has a significant ability to determine priority zones for the rehabilitation plan, by identifying multiple clusters and their according levels. Therefore, the proposed method can be used as a practical and flexible tool to prioritize the rehabilitation process of water networks and identify the main causes of failure events, avoiding complex computational methods or the personal judgment of experts.

**Keywords:** Density-Based Clustering, Spatial Analysis, Infrastructure Networks, Failure, Condition Assessment, Water Distribution Network, OPTICS Algorithm.

Received: October 28, 2021

Accepted: February 6, 2022

## خوشه‌بندی مکانی حوادث شبکه‌های توزیع آب برای اولویت‌بندی نواحی بازسازی

امیرمحمد آریایی<sup>۱</sup>، محمدرضا جلیلی قاضی‌زاده<sup>۲\*</sup>  
و ایمان مصلحی<sup>۳</sup>

### چکیده

شبکه‌های آبرسانی ارتباط مستقیمی با توسعه اقتصادی و اجتماعی جوامع شهری دارند. با توجه به افزایش سریع جمعیت و توسعه شهرنشینی لازم است تا این شبکه‌ها به طور مناسب نگهداری و بازسازی شوند. در این مقاله، با بهره‌گیری از روش یادگیری غیرنظارتی و با استفاده از الگوریتم خوشه‌بندی مبتنی بر چگالی OPTICS مناطق اولویت‌دار شبکه‌های توزیع آب جهت بازسازی تعیین گردیدند. ۳۶۱ حادثه شکستگی رخ داده در یکی از پهنه‌های شبکه توزیع آب مشهد مورد بررسی واقع شد. این الگوریتم در مجموع ۱۶ خوشه را در پنج سطح متفاوت اهمیت شناسایی کرد. این ۱۶ خوشه که برخی زیرخوشه دیگری محسوب می‌شوند از منظر نرخ شکست با یکدیگر مورد مقایسه قرار گرفتند و سه خوشه که به ترتیب نرخ شکستگی برابر با ۷۹/۱، ۶۲/۲ و ۵۰/۱ (حادثه/۱۰۰ کیلومتر/سال) را داشتند، به‌عنوان اولویت‌های اصلی بازسازی معرفی شدند. نرخ شکست میانگین در کل شبکه مورد بررسی برابر با ۱۴/۸ بود. در ادامه، خوشه‌های شناسایی شده از منظر جنس و قطر لوله نیز مورد ارزیابی قرار گرفته و عوامل مؤثر در نرخ شکستگی آن‌ها، بررسی شدند. نتایج نشان داد که الگوریتم خوشه‌بندی OPTICS با شناسایی خوشه‌های متعدد و سطح‌بندی آن‌ها توانایی قابل‌توجهی در تشخیص پهنه‌های اولویت‌دار برای برنامه‌های بازسازی دارد؛ بنابراین از روش پیشنهادی به‌عنوان ابزاری کاربردی و انعطاف‌پذیر برای اولویت‌بندی بازسازی شبکه توزیع آب، شناسایی علل اصلی حوادث با پرهیز از روش‌های محاسباتی پیچیده و طولانی و به دور از قضاوت شخصی کارشناسان می‌توان استفاده کرد.

**کلمات کلیدی:** خوشه‌بندی مبتنی بر چگالی، تحلیل مکانی، شبکه‌های زیرساختی، شکستگی، ارزیابی وضعیت، شبکه توزیع آب، الگوریتم OPTICS.

تاریخ دریافت مقاله: ۱۴۰۰/۸/۶

تاریخ پذیرش مقاله: ۱۴۰۰/۱۱/۱۷

1- M.Sc. Student, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University. Email: a.aryayee@mail.sbu.ac.ir

2- Associate Professor, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University. Email: m\_jalili@sbu.ac.ir

3- Ph.D. Student, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University. Email: i\_moslehi@sbu.ac.ir

\*- Corresponding Author

Dor: [20.1001.1.17352347.1400.17.4.10.7](https://doi.org/10.17352/347.1400.17.4.10.7)

۱- دانشجوی کارشناسی ارشد مهندسی عمران- محیط‌زیست، دانشکده عمران، آب و محیط‌زیست، دانشگاه شهید بهشتی.

۲- دانشیار، دانشکده عمران، آب و محیط‌زیست، دانشگاه شهید بهشتی.

۳- دانشجوی دکتری مهندسی عمران- آب، دانشکده عمران، آب و محیط‌زیست، دانشگاه شهید بهشتی.

\*- نویسنده مسئول

بحث و مناظره (Discussion) در مورد این مقاله تا پایان بهار ۱۴۰۱ امکان‌پذیر است.



این مدل‌ها در صورت وجود داده‌های مطلوب است (Kleiner and Rajani, 2010; Xu et al., 2011; Kabir et al., 2015).

۲) مدل‌های مدیریت دارایی (AMMs)<sup>۴</sup> که به‌عنوان یک رویکرد عملی، از توسعه مدل‌های شکست حاصل می‌گردند. این مدل‌ها توانایی قابل توجهی در تعیین شرایط مکانیکی شبکه‌ها، تجزیه و تحلیل طول عمر آن‌ها و همچنین ارائه راهکارهای بهینه بازسازی و نوسازی شبکه‌های آبرسانی دارند (Salehi et al., 2018). با این حال، مطالعات اخیر نشان داده است که قابلیت این مدل‌ها در تهیه و تنظیم برنامه‌های منسجم برای بازسازی و نوسازی شبکه‌های توزیع آب با عدم قطعیت و محدودیت‌هایی همراه است (Kleiner and Rajani, 2001; Engelhardt et al., 2003; Scheidegger et al., 2015).

۳) مدل‌های تصمیم‌گیری (DMMs)<sup>۵</sup>، که با عنوان مدل‌های اولویت‌بندی و بهینه‌سازی جهت اجرای فرآیند نوسازی در شبکه‌های آبرسانی و همچنین بر اساس الگوهای تصمیم‌گیری توسعه می‌یابند (Salehi et al., 2018). در بیان کلی، این مدل‌ها بر اساس رویکردهای تک یا چند معیاره کارکرد دارند. در رویکرد اول، تصمیم‌گیری بر اساس یک معیار حاصل می‌شود و این در حالی است که در روش دوم، چندین معیار و ضابطه وجود دارد (Tzeng and Huang, 2011). با توجه مجموعه مطالعات انجام شده، می‌توان اذعان داشت که کاربرد مدل‌های تصمیم‌گیری‌های تک معیاره در حل مسائل مرتبط با شبکه‌های آبرسانی به صورت محدودی صورت پذیرفته (Gomes et al., 2008) و در مقابل، استفاده از مدل‌های چند معیاره در این زمینه، به صورت گسترده‌ای مدنظر بوده است (Roobzahani et al., 2012). مدل‌های تصمیم‌گیری چند معیاره به دو زیرگروه اصلی تقسیم می‌شوند که رویکرد حاکم بر هر یک از آن‌ها عبارت است از تصمیم‌گیری چندهدفه (MODM) و تصمیم‌گیری چند شاخصه (MADM)؛ گروه اول مبتنی بر روش‌های بهینه‌سازی چند معیاره هستند و در مقابل، دسته دوم از این روش‌ها بر پایه تکنیک‌های تحلیلی چندمعیاره توسعه می‌یابند (Tzeng and Huang, 2011). مطالعات اخیر در زمینه استفاده از مدل‌های تصمیم‌گیری چند معیاره نشان می‌دهد که انواع چند شاخصه آن کارایی بهتری را در قیاس با مدل‌های چندهدفه و در ارتباط با مسائل بهسازی و نوسازی شبکه‌های توزیع آب دارا می‌باشند (Scholten et al., 2014). این دسته از روش‌ها قادر است تا عدم قطعیت مرتبط با متغیرهای تصمیم‌گیری را حل نماید و بر همین مبنا در چنین حالتی، مجموعه از جواب‌های بهینه به عنوان راه‌حل‌های مسئله تولید شده که در قیاس با وجود معیارهای مختلف، از اثرگذاری بیشتری در حل مسئله برخوردار است. افزون بر موارد یاد شده، مدل‌های چند معیاره با مسائلی که به انتخاب از میان

سامانه‌های زیرساختی آب شرب، نقش اساسی در توسعه اقتصادی- اجتماعی جوامع مدرن و حفاظت از سلامت عمومی داشته (Nazemi and Madani, 2018) و در حالت کلی، هدف از این سامانه‌ها به‌خصوص شبکه‌های توزیع آب (WDNs)<sup>۱</sup>، فراهم کردن دسترسی قابل اطمینان به آب سالم برای برآوردن کلیه نیازهای مصرفی شهری، صنعتی و آتش‌نشانی یک جامعه است. در همین راستا و با توجه به رشد سریع جمعیت و افزایش شهرنشینی و مقادیر تقاضا، شبکه‌های توزیع آب باید به‌طور مناسب نگهداری و به‌روز شوند. در سال‌های اخیر، طیف گسترده‌ای از تحقیقات در زمینه تحلیل اکتشافی<sup>۲</sup> و پیش‌گویانه از طریق بررسی داده‌های حوادث لوله‌های شبکه برای مدیریت دارایی زیرساخت‌ها، انجام شده است. به منظور بازسازی و نوسازی سیستم‌های آبرسانی، همواره صرف هزینه‌های گزافی لازم است تا اطمینان حاصل شود که زیرساخت‌های خط لوله به عملکرد خوب خود ادامه می‌دهند (Baird, 2010). از آنجایی که بسیاری از شرکت‌های آب و فاضلاب، بودجه محدودی را برای این منظور در اختیار دارند، داشتن یک چارچوب مؤثر در مدیریت دارایی‌ها امری حیاتی است. به عبارت دیگر، این موضوع را اطمینان می‌بخشد که سرمایه‌گذاری‌های لازم در آسیب‌پذیرترین مناطق انجام شود و امکانی فراهم گردد که خطرات ناشی از شکستگی لوله‌ها به‌طور مناسب کاهش یابند.

طی دهه‌های اخیر، تحقیقات زیادی در زمینه برنامه‌های بازسازی و نوسازی شبکه‌های توزیع آب صورت پذیرفته است که مدل‌های توسعه داده شده در این تحقیقات را در سه گروه اصلی می‌توان دسته‌بندی کرد:

۱) مدل‌های شکست (FMs)<sup>۳</sup>، که بر شناخت الگوی حاکم بر رخداد حوادث و متغیرهای مربوط به آن متمرکز بوده و تمام عوامل مؤثر در خرابی (پارامترهای عملیاتی، محیطی و مشخصات فیزیکی و غیره) و تعدد آن‌ها را تعیین می‌نمایند. این مدل‌ها را به مدل‌های فیزیکی، مدل‌های آماری و مدل‌های مبتنی بر داده‌کاوی می‌توان طبقه‌بندی کرد (Xu et al., 2011). نکته قابل توجه در این زمینه آن است که شرایط دو لوله کاملاً با یکدیگر یکسان نیستند، حتی اگر از یک جنس و اندازه ساخته شده باشند، زیرا عوامل غیرساختاری مانند آب و هوا، بارهای خارجی و غیره می‌تواند متفاوت باشد. در عین حال، انجام تجزیه و تحلیل آماری در مورد الگوی شکست یک لوله امری غیرمنطقی است؛ زیرا اطلاعات کافی جهت تحلیل معتبر برای این لوله وجود ندارد (Salehi et al., 2018). با این وجود در سال‌های گذشته مطالعاتی در این زمینه‌ها صورت گرفته است که حاکی از توانایی بالای

مورنوویل (ایالت پنسیلوانیا) و یافتن الگوهای مناسب جهت پیش‌بینی آن استفاده نمودند. در این مطالعه، حوادث موجود از طریق دو الگوریتم مذکور خوشه‌بندی شده که مبتنی بر آن، الگوریتم DBSCAN تطابق بیشتری با معیارهای خوشه‌بندی از خود نشان داد. Abokifa and Sela (2019) با استفاده از آنالیز خودهمبستگی مکانی (SAA)<sup>۶</sup> ابتدا حضور و اهمیت آماری الگوهای مکانی در داده‌ها را بر اساس SAA جهانی مشخص نموده و سپس بر اساس SAA محلی، انواع مختلفی از خوشه‌ها (به‌عنوان مثال، نقاط پراهمیت و کم‌اهمیت) و مکان آن‌ها را در شبکه مشخص کردند. جهت ارتباط برقرار کردن میان شبکه جهانی و محلی در SAA نیز، روشی با هدف تعیین وزن‌های مکانی تعریف شده است. در مرحله پایانی نیز، SAA برای شناسایی سلسله مراتب خوشه‌بندی تکرار شده و با شناسایی نقاط مهم و کم‌اهمیت از حوادث، وابستگی آن‌ها به ویژگی‌های شبکه توزیع آب (به‌عنوان مثال، لوله و سن) بررسی گردیدند.

Aryayee et al. (2020) با استفاده از داده‌های حوادث شکستگی در منطقه‌ای از شبکه توزیع آب مشهد و با به‌کارگیری الگوریتم DBSCAN برای شناسایی مناطق با بالاترین اولویت، جهت بازسازی و نوسازی شبکه توزیع آب استفاده کردند. در این رویکرد، به‌منظور بهینه‌سازی پارامترهای الگوریتم DBSCAN شامل شعاع همسایگی و حداقل نقاط همسایگی، گراف K-امین همسایگی مورد استفاده قرار گرفت. نتایج نشان داد که این روش، توانایی شناسایی مناطق با اولویت بالاتر در امر بازسازی و نوسازی را دارا بوده و از این متدولوژی می‌توان در تحلیل و اولویت‌بندی بازسازی شبکه‌های توزیع آب استفاده کرد.

یکی از ویژگی‌های الگوریتم OPTICS، شناسایی خوشه‌هایی است که چگالی متفاوتی دارند (Ankerst et al., 1999). بنابراین در مقاله حاضر از این روش جهت تحلیل مکانی حوادث شکستگی در شبکه توزیع آب بهره برده شده این الگوریتم خوشه‌هایی را در سطوح مختلف سلسله مراتبی شناسایی می‌کند، بدین صورت که خوشه‌های سطوح اولیه به چند زیرخوشه کوچک‌تر تقسیم می‌گردند. با این عمل، احتمال شناسایی خوشه‌هایی که نرخ شکست بالایی دارند، افزایش می‌یابد که این موضوع از منظر سرمایه‌گذاری با توجه به محدودیت‌های مالی شرکت‌های آب و فاضلاب با اهمیت است. علاوه بر این، با توجه به این فرض که چگالی بالای حوادث شکستگی ناشی از عامل زمینه‌ای مانند نوسانات فشار، عمر شبکه و غیره است، یافتن خوشه‌هایی که با محدوده جغرافیایی کوچک‌تر نرخ شکست بالاتری دارند، شناسایی علل اصلی رخداد حوادث را نیز افزایش می‌دهد. بنابراین خوشه‌هایی که در سطوح بالاتری قرار می‌گیرند، از منظر ویژگی‌های قطر و جنس و غیره نرخ شکست بیشتری را نیز به خود اختصاص

چندین گزینه می‌پردازند، انطباق بیشتری دارند و این در حالی است که مدل‌های چندهدفه با فرآیند طراحی مرتبط هستند (Tzeng and Huang, 2011). بر همین اساس و با توجه به این که پروژه‌های نوسازی و بهسازی شبکه‌های آب اساساً با انتخاب بهترین گزینه از میان مجموعه راه‌حل‌های ممکن همراه هستند و به این واسطه می‌کوشند تا کیفیت آبرسانی به مشترکین را بهبود بخشند، کارایی مدل‌های چندمعیاره در قیاس با مدل‌های چندهدفه با بهره‌وری بالاتری ارزیابی می‌گردد (AWWA, 2014). تحقیقات اخیر نشان داده که مدل‌های چند شاخصه عموماً در مطالعاتی با هدف اجرای بهسازی و نوسازی شبکه‌های توزیع آب به صورت گسترده به کار گرفته شده‌اند که غالباً با اجرای اولویت‌بندی نیز همراه بوده‌اند (Tschekner-Gratl et al., 2016).

یکی دیگر از شاخه‌های اصلی مدل‌های تصمیم‌گیری، مدل‌های تصمیم‌گیری مبتنی بر خوشه‌بندی فضایی است. طی سال‌های اخیر، یادگیری ماشین که یکی از شاخه‌های اصلی داده‌کاوی است به‌عنوان ابزاری بسیار کارا جهت پیش‌بینی حوادث در لوله‌ها مطرح شده است. مدل‌های یادگیری ماشین نظارت نشده قادر به شناسایی و مدل‌سازی روابط پیچیده‌ای هستند که بین متغیرهای پیش‌بینی و شکستگی لوله برقرار است. در این راستا، تحلیل خوشه‌بندی فضایی که از جمله ابزار مهم داده‌کاوی است، به‌صورت گسترده مدنظر قرار گرفته و کاربرد آن در زمینه‌هایی نظیر طبقه‌بندی داده‌ها در تکنولوژی‌های مهندسی، تفکیک داده‌های پیوسته، ردیابی نقطه خطا، پالایش داده‌ها از نقطه‌نظر اختلال و نویز، شناسایی الگوها و همچنین پردازش تصاویر قابل توجه می‌باشد (Sattar et al., 2019). امروزه با توسعه مبانی علم سامانه اطلاعات جغرافیایی (GIS) و کاربرد روزافزون سنسورهای پیش آنالین این امکان فراهم آمده تا با توجه به در اختیار داشتن اطلاعات لحظه‌ای شبکه، شناسایی پهنه‌های با ریسک بالا امکان یافته و از صرف هزینه‌های غیر ضروری که ناشی از قضاوت شخصی کارشناسان بوده جلوگیری شود. نکته دیگری که در این زمینه حائز اهمیت است، فراوانی داده‌های پیش غیرمستقیم بوده که امر پردازش اطلاعات را با چالش‌هایی مواجه می‌سازد. در این راستا، مبانی داده‌کاوی می‌تواند با هدف آشکار نمودن قواعد موجود در داده‌های پیش غیرمستقیم و به‌خصوص تحلیل داده‌های نقطه‌ای تصادفی مورد کاربرد قرار گیرد. Ester et al. (1996) روش تحلیل خوشه‌بندی مبتنی بر چگالی را معرفی نمود و در همین راستا، قوانین لازم جهت تشریح دقیق الگوریتم خوشه‌بندی فضایی مبتنی بر چگالی با کاربردهای همراه با اختلال (DBSCAN)<sup>۶</sup> تبیین گردیده است. Oliveira et al. (2009) از الگوریتم‌های DBSCAN و EM<sup>۷</sup> برای ارزیابی اثربخشی استفاده از روش‌های خوشه‌بندی در تجزیه و تحلیل حوادث شبکه آب شهر

(شکل ۱). برای تعریف شیء مرزی نیاز است تا دو مشخصه دیگر نیز تعریف گردند، (۱) دسترسی پذیر چگالی مستقیم، (۲) دسترسی پذیر چگالی. مطابق با شکل ۲-الف، شیء  $p$  دسترسی پذیر چگالی مستقیم از شیء  $q$  است، اگر اولاً  $p$  جزء همسایه‌های شیء  $q$  باشد و ثانیاً شیء  $q$  یک شیء مرکزی باشد. همچنین، شیء  $p$  دسترسی پذیر چگالی از شیء  $q$  است، اگر مطابق با شکل ۲-ب یک زنجیره از نقاط  $p_0, p_1, p_2, \dots, p_n$  وجود داشته باشد؛ درحالی که  $p_0 = q$  و  $p_n = p$ ، به گونه‌ای که هر نقطه مانند  $P_{i+1}$  دسترسی پذیر چگالی مستقیم از نقطه قبلی ( $P_i$ ) باشد. با توجه به تعاریف گفته شده، شیئی که شیء مرکزی نباشد، منتها از یک شیء مرکزی دیگر دسترسی پذیر چگالی باشد، شیء مرزی نام دارد (شکل ۲-ب) (Ester et al., 1996).

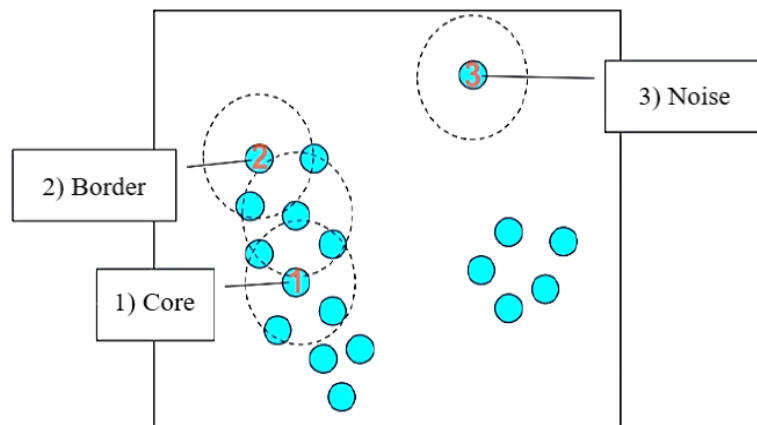
با این حال یکی از معایب الگوریتم DBSCAN، عدم قابلیت تشخیص خوشه‌های با چگالی متفاوت است (Ankerst et al., 1999). این در حالی است که یکی از خصوصیات مهم مجموعه داده‌های دنیای واقعی این است که خوشه‌های موجود در این مجموعه داده‌ها، وجود چگالی‌های محلی متفاوت است. بنابراین، Ankerst et al. (1999) الگوریتم OPTICS را معرفی کردند که نوع تعمیم‌یافته‌ای از الگوریتم DBSCAN است. این الگوریتم خوشه‌بندی را با روشی سلسله‌مراتبی و همراه با ایجاد خوشه‌های تودرتو انجام می‌دهد (شکل ۳). در این حالت، خوشه‌ها در سطوح مختلفی تشکیل می‌شوند؛ به طوری که زیر خوشه‌هایی که در خوشه مستقل دیگری جای گرفته‌اند، جزء لایه‌های بعدی خوشه‌بندی مدنظر قرار می‌گیرند.

می‌دهند تا امکان تعیین نواحی اولویت‌دار جهت بازسازی و نوسازی شبکه آسان‌تر گردد. به این ترتیب که در گام اول، داده‌های مرتبط با حوادث در لوله‌های شبکه آبرسانی گردآوری و پس از پیش‌پردازش داده‌ها، پارامترهای بهینه مرتبط با الگوریتم خوشه‌بندی مبتنی بر چگالی تعیین شده و نهایتاً با اجرای الگوریتم یادشده و تحلیل خوشه‌های حاصل، اولویت‌بندی نواحی موردتوجه قرار می‌گیرد. همچنین به منظور ارزیابی شبکه معرفی شده در این تحقیق، پیاده‌سازی آن برای یک پهنه بهره‌برداری واقع در شبکه آبرسانی شهر مشهد مورد انجام شده است.

## ۲- روش انجام کار

### • خوشه‌بندی مبتنی بر چگالی

یکی از متداول‌ترین الگوریتم‌های خوشه‌بندی مبتنی بر چگالی، الگوریتم DBSCAN است که بر مبنای دو پارامتر شعاع همسایگی ( $\epsilon$ ) و حداقل نقاط همسایگی (MinPts) می‌باشد که در واقع حداقل تعداد اشیایی است که باید در محدوده  $\epsilon$  باشد تا یک خوشه شکل بگیرد (Ester et al., 1996). همچنین، اگر در محدوده شعاع همسایگی یک شیء، به تعداد حداقل نقاط همسایگی شیء وجود داشته باشد؛ آنگاه به آن شیء مرکزی گفته می‌شود (شکل ۱). در این حالت، اگر اشیایی در مجموعه داده باقی بمانند که متعلق به هیچ خوشه‌ای نباشند و یا در محدوده  $\epsilon$  اطراف خود نیز به اندازه MinPts همسایه نداشته باشند؛ آنگاه به آن اشیا داده پرت (Noise) گفته می‌شود



$$\text{MinPts} = 5$$

Fig. 1- Definition of the points in DBSCAN algorithm, 1- Core point, 2- Border point, and 3-Noise for MinPts= 5

شکل ۱- تعریف نقاط در الگوریتم DBSCAN، ۱- شیء مرکزی، ۲- شیء مرزی، ۳- داده پرت برای  $\text{MinPts} = 5$  (Moslehi et al., 2019)

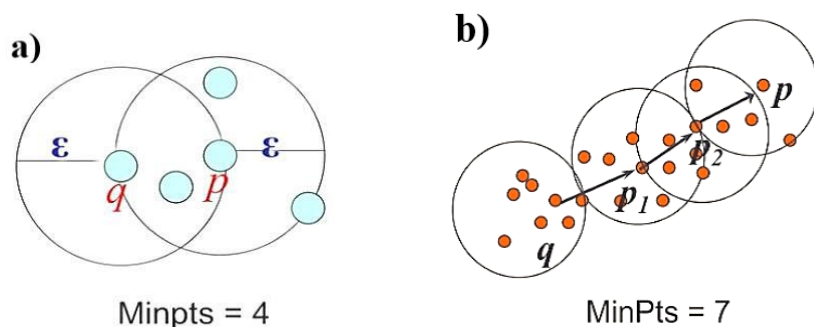


Fig. 2- Graphical presentation of the relationship between three points in DBSCAN algorithm, a) Directly density reachable for MinPts= 4, b) density reachable for MinPts= 7

شکل ۲- نمایش گرافیکی روابط میان سه داده در الگوریتم DBSCAN، الف) دسترسی پذیری چگالی مستقیم برای  $MinPts = 4$ ، ب) دسترسی پذیری چگالی برای  $MinPts = 7$  (Moslehi et al., 2019)

کوچک‌ترین فاصله‌ای که شیء P دسترسی‌پذیر چگالی مستقیم از شیء O باشد؛ به شرطی که O یک شیء مرکزی باشد، فاصله دسترسی‌پذیری  $(RD_P)$  گفته می‌شود. در این حالت، فاصله دسترسی‌پذیری نمی‌تواند کوچک‌تر از فاصله مرکز شیء O باشد؛ زیرا برای فواصل کمتر هیچ شیئی به‌طور مستقیم از O قابل دسترسی به تراکم نیست (Ankerst et al., 1999). به بیانی دیگر، فاصله دسترسی‌پذیری شیء P برابر است با حداکثر مقدار بین این دو فاصله:

(۱) فاصله اقلیدسی بین دو شیء P و O (نزدیک‌ترین همسایه آن) که  $d_{PO}$  نام دارد،

(۲) فاصله اقلیدسی بین شیء P و  $MinPts-1$  - اُمین نزدیک‌ترین همسایه آن که فاصله مرکز (CD) نام دارد.

$$RD_P = \max(d_{PO}, CD_P) \quad (1)$$

#### • سازوکار الگوریتم OPTICS

الگوریتم OPTICS خوشه‌ها را به کمک یک گراف ترتیبی شناسایی می‌کند. اگر مقادیر فاصله دسترسی‌پذیری برای هر شیء به ترتیب ترسیم شده باشد، می‌توان خوشه‌ها و زیرخوشه‌های موجود در مجموعه داده را به‌صورت بصری درک کرد. در این نمودار، خوشه‌ها در مناطق مقعر نمودار دسترسی‌پذیری جای دارند (شکل ۴-الف). در واقع می‌توان گفت که پس از محاسبه فاصله دسترسی‌پذیری (RD) همه داده‌ها، حداقل تعداد نقاط همسایگی ( $MinPts$ ) و حداقل شیب لازم جهت تفکیک خوشه‌ها ( $\xi$ ) تعیین می‌گردد. مقدار  $\xi$  را Ankerst et al. (1999) برابر با  $0.5$  و De Oliveira et al. (2011) برابر با  $0.1$  در نظر گرفته‌اند. پس از ترسیم نمودار فاصله دسترسی‌پذیری، شیب تغییرات بین هر RD نسبت به RD ماقبل خود در نمودار دسترسی‌پذیری بررسی شده و چنانچه از مقدار معینی تجاوز کند، خوشه از همان نقطه شروع می‌شود (Reach Start). همچنین، هنگامی که

چنانچه الگوریتم بتواند سطوح بیشتری را شناسایی کند و هرچه قدر داده‌های موجود در سطوح آخر نیز بیشتر باشد، احتمال یافتن خوشه‌های همگن‌تر نیز افزایش پیدا می‌کند که این موضوع هدف اصلی خوشه‌بندی است.

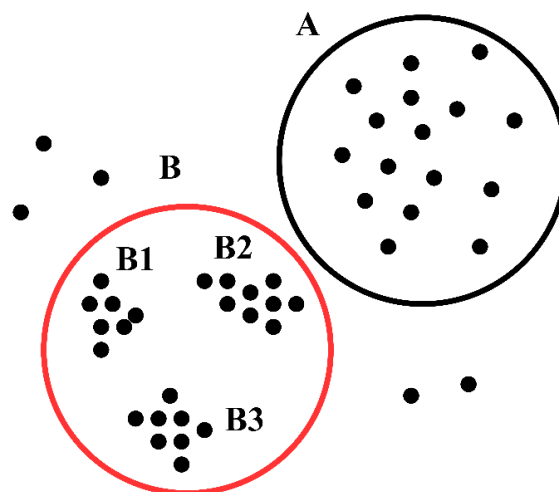


Fig. 3- Clusters with different density  
شکل ۳- خوشه‌های با چگالی متفاوت (Ankerst et al., 1999)

الگوریتم OPTICS به‌طور صریح خوشه‌های یک مجموعه داده را تولید نمی‌کند، بلکه در عوض یک گراف ترتیبی از پایگاه داده ایجاد می‌کند. در واقع به کمک این گراف، طی فرایند خوشه‌بندی پارامتر شعاع همسایگی ( $\epsilon$ ) متغیر است تا خوشه با هر شکل و اندازه‌ای را شناسایی کند. همچنین این ویژگی سبب می‌شود تا الگوریتم OPTICS خوشه‌هایی را به‌صورت تودرتو و سلسله‌مراتبی تشخیص دهد. دو مؤلفه مهم در الگوریتم OPTICS عبارت‌اند از: (۱) فاصله مرکز و (۲) فاصله دسترسی‌پذیری. فاصله مرکز شیء P یا  $(CD_P)$ ، به معنی کوتاه‌ترین فاصله  $(\epsilon)$  بین شیء P و  $MinPts-1$  - اُمین همسایه‌اش است، به شرطی که P یک شیء مرکزی باشد. همچنین،

روندنمای روش پیشنهادی را نشان می‌دهد که در انتها و پس از تشکیل خوشه‌ها می‌توان با ارزیابی آن‌ها و تفسیر نتایج آن به اطلاعات مورد نیاز از مجموعه داده‌های ورودی دست یافت.

RD برابر با نقطه (Reach Start = Reach End) شود، خوشه خاتمه می‌یابد (شکل ۴-ب). تا زمانی که خوشه‌ای به نقطه (Reach End) نرسیده است، چنانچه باز هم تغییرات شیب از آستانه تعریف شده ( $\xi$ ) شود، زیرخوشه‌های دیگری تشکیل می‌گردد (شکل ۴-الف). شکل ۵

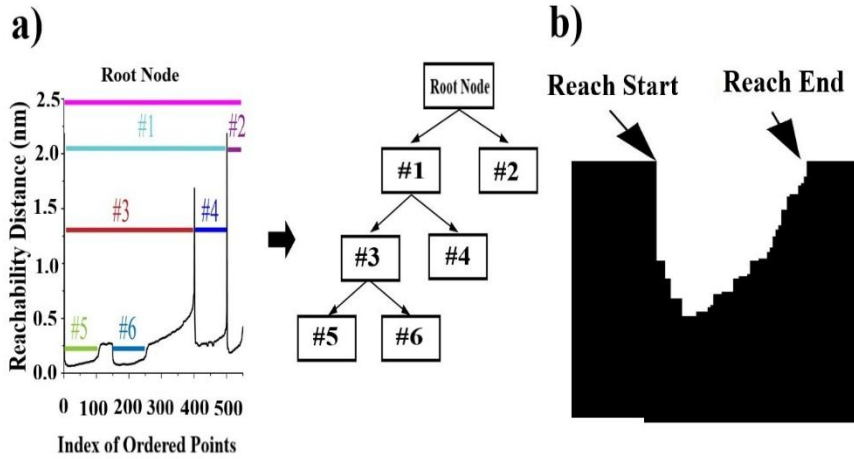


Fig. 4- a) Reachability-plot for clusters & sub-clusters identified by the OPTICS algorithm, b) Start and end points of identified clusters based on OPTICS algorithm

شکل ۴- الف) نمودار دسترسی پذیری برای خوشه‌ها و زیرخوشه‌های شناخته شده توسط الگوریتم OPTICS (Wang et al., 2019)، ب) نقاط شروع و خاتمه خوشه‌های شناسایی شده بر مبنای الگوریتم OPTICS (Ankerst et al., 1999)

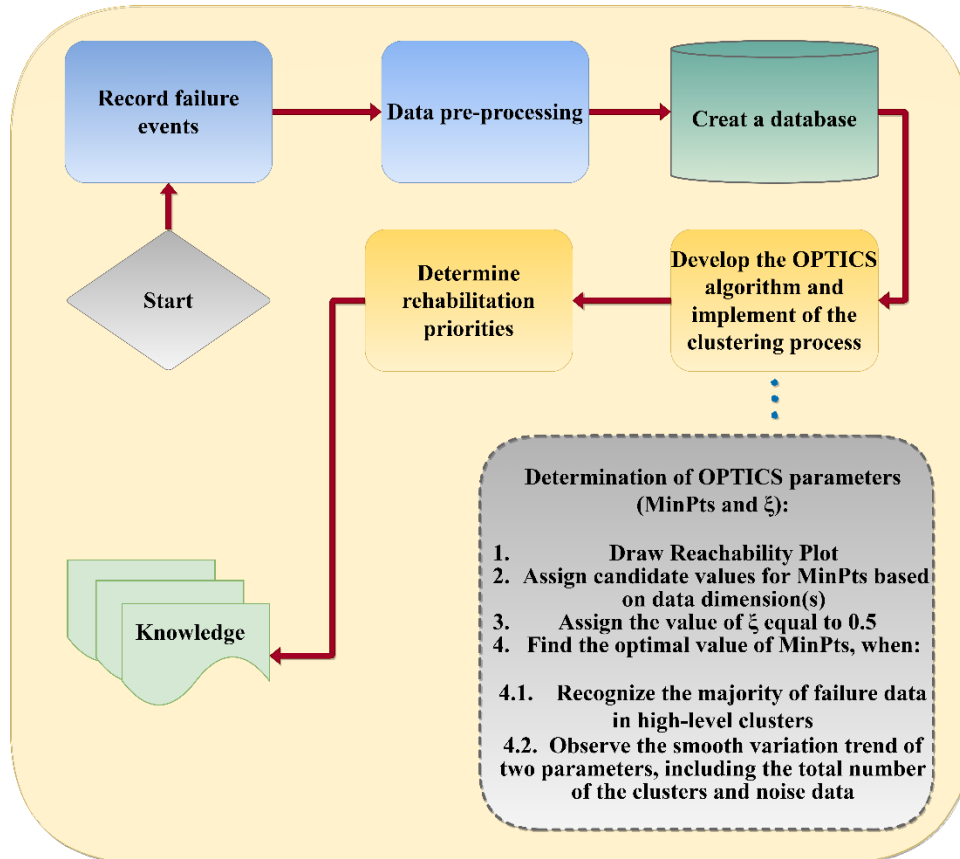


Fig. 5- Flow diagram for the proposed methodology

شکل ۵- روندنمای روش پیشنهادی

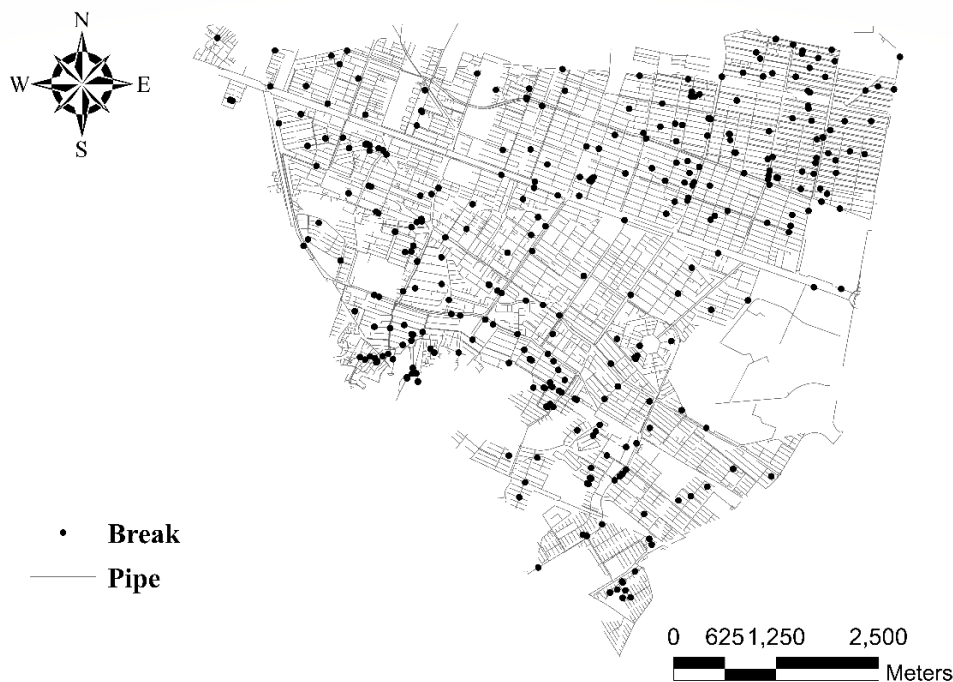
### ۳- مطالعه موردی

۶). در پژوهش حاضر، داده‌های قطر و جنس لوله مورد بررسی واقع شده است. همچنین، خوشه‌بندی مجموعه داده‌ها بر اساس دو پارامتر طول و عرض جغرافیایی صورت گرفته است و خوشه‌های تشکیل شده از منظر قطر، جنس و نرخ شکست حوادث (تعداد/ ۱۰۰ کیلومتر/ سال) مورد بررسی بیشتر واقع شده‌اند. جدول ۱، مشخصات لوله‌های به کار گرفته شده در شبکه اعم از جنس، طول، قطر، تعداد حوادث و نرخ شکست را نشان می‌دهد.

محدوده مورد مطالعه در این تحقیق، یک منطقه از شرکت آب و فاضلاب شهر مشهد است. طول شبکه توزیع در این منطقه حدود ۷۲۴ کیلومتر است. بازه زمانی مورد بررسی در این تحقیق از بهمن‌ماه سال ۱۳۹۴ لغایت مهرماه ۱۳۹۸ بوده که در این بازه، تعداد ۳۶۱ حادثه شکستگی که منجر به قطعی جریان آب در شبکه توزیع گردیده و جزییات آن نیز همراه با مختصات مکانی به ثبت رسیده است (شکل

**Table 1- Summary of collected dataset**  
جدول ۱- خلاصه‌ای از مجموعه داده‌های گردآوری شده

Pipe Material	Length (km)	Diameter (mm)	Failure No.	Failure Rate (No./100 km/year)
ACC	119.482	80	124	28.65
	202.339	100-200	58	7.91
	62.981	200-300	10	4.38
PE	202.594	100 <	129	17.58
	15.819	100-200	9	15.71
PVC	0.926	100 <	2	59.63
	0.522	100 <	12	634.71
	2.774	100-200	3	29.86
ST	2.249	200-300	1	12.28
	6.649	300	1	4.15
	1.113	600	1	24.81
	1.753	1000	5	78.75
	0.706	100 <	3	117.32
CI	0.509	200-300	1	54.24
	1.5	300-400	2	36.81



**Fig. 6- The recorded pipe breaks in water distribution network of the study area**  
شکل ۶- نقاط حوادث شکستگی ثبت شده در شبکه توزیع آب منطقه مورد مطالعه

## • تعیین پارامترهای الگوریتم OPTICS

پارامترهای تأثیرگذار در الگوریتم OPTICS، دو مؤلفه حداقل شیب لازم در نمودار فاصله دسترسی‌پذیری جهت تفکیک خوشه‌ها ( $\xi$ ) و حداقل نقاط همسایگی (MinPts) می‌باشند. در این راستا، نخست تعیین MinPts بهینه مورد توجه قرار گرفته است. بر این مبنای، از پارامتر تعداد ابعاد داده‌های حوادث کمک گرفته شده (Sander et al., 1998) و با توجه به این که داده‌های موجود مکانی بوده و از دو بعد طول و عرض جغرافیایی تشکیل شده‌اند، بنابراین از ۲، ۳، ۴، ۵ و ۶ برابر ابعاد داده‌ها استفاده گردیده است. این امر بدان معنا بوده که برای اجرای فرآیند خوشه‌بندی، حداقل نقاط همسایگی برابر با ۴، ۶، ۸، ۱۰ و ۱۲ در نظر گرفته شده است.

همانطور که در بخش روش انجام کار توضیح داده شد، برای تشخیص خوشه‌هایی که چگالی متفاوتی دارند از پارامتر حداقل شیب ( $\xi$ ) در الگوریتم OPTICS بهره برده شده است. شعاع همسایگی در این حالت می‌تواند تا  $(\pm \%)$  متغیر باشد. از آنجایی که پارامترهای ورودی الگوریتم، مستقیماً با نتایج خروجی و چگالی خوشه‌ها ارتباط دارد، اگر  $\xi$  خیلی بزرگ انتخاب شود احتمال می‌رود تا داده‌هایی را که ذاتاً به خوشه‌ای دیگر تعلق دارند، به آن خوشه متصل گردند. این امر تحلیل خوشه‌ها و شناسایی عوامل زمینه‌ای حوادث را دچار اختلال می‌کند. از طرفی دیگر، اگر  $\xi$  خیلی کوچک انتخاب شود، دیگر خوشه‌هایی با چگالی متفاوتی در مجموعه داده شناسایی نخواهد شد. از این رو در این تحقیق و بنا به توصیه (Ankerst et al. (1999)، مقدار  $\xi$  برابر با ۰/۰۵ در نظر گرفته شده است.

خوشه‌ها از  $MinPts=8$  به بعد، دچار تغییرات چشم‌گیری نشده و تقریباً در تعداد ۱۵ خوشه همگرا می‌شود. البته ذکر این نکته ضروری است که چنانچه حداقل نقاط همسایگی به بیش از ۱۲ نقطه افزایش یابد، قطعاً از جایی به بعد تعداد خوشه‌ها به دلیل تغییرات چگالی، دستخوش تغییرات خواهند شد. اما از آنجایی که باید چگالی خوشه‌ها متناسب با مجموعه داده مورد بررسی انتخاب شود، از محاسبات بیشتر که تنها خطا را افزایش می‌دهد، اجتناب شد. در مجموع با توجه به تحلیل صورت گرفته در این بخش، می‌توان نتیجه گرفت که الگوریتم OPTICS در  $\xi = 0.05$  به ازای حداقل نقاط همسایگی معادل ۸، ۱۰ و ۱۲ حساسیت کمتری از خود نسبت به تغییرات MinPts نشان داده و به نظر می‌رسد که پارامتر MinPts بهینه یکی از این سه مقدار است. بنابراین در ادامه به بررسی بیشتر این پارامتر پرداخته می‌شود. همچنین، در این نمودار، تعداد داده‌های پرت نیز نشان داده شده است. بر اساس این نمودار، تعداد داده‌های پرت شناسایی شده توسط الگوریتم OPTICS برای  $\xi = 0.05$ ، با افزایش حداقل نقاط همسایگی هیچ تغییری نداشته و به صورت ثابت باقی مانده است. از مقایسه تعداد داده‌های پرت در شکل ۷، این موضوع برداشت می‌شود که تغییرات پارامتر حداقل نقاط همسایگی تأثیر خاصی بر نتایج الگوریتم OPTICS نداشته و الگوریتم به راحتی این مقدار تقریباً یکسان از داده‌های پرت را در حالات مختلف شناسایی می‌کند. لازم به ذکر است که داده‌های پرت ممکن است به دلیل خطای GPS رخ داده باشند و یا در اطراف این حوادث، نشت گزارش نشده وجود داشته باشد. بنابراین تا زمانی که تراکم این حوادث کم باشد نمی‌توانند گزینه‌های مطلوبی برای بازسازی و نوسازی باشند.

## • تحلیل طبقه‌بندی خوشه‌ها

در شکل ۸، تعداد داده‌های موجود در سطوح مختلف خوشه‌بندی که در نتیجه اجرای الگوریتم OPTICS با مقادیر مختلف حداقل نقاط همسایگی معادل ۸، ۱۰ و ۱۲ در  $\xi = 0.05$  ایجاد شده‌اند، محاسبه و نتایج به نمایش درآمده است. بر اساس این نمودار، مشاهده می‌شود که الگوریتم OPTICS به ازای مقادیر مختلف حداقل نقاط همسایگی، به یک اندازه داده در لایه اول شناسایی کرده است. همچنین، مشخص است که در لایه‌های بعدی با افزایش MinPts، تعداد داده‌ها ابتدا افزایش و سپس کاهش می‌یابد. بر این مبنای می‌توان ادعان داشت که چون الگوریتم OPTICS توانسته در  $MinPts=10$  داده‌های بیشتری را در سطوح بالاتر شناسایی کند، در این MinPts، چگالی مطلوب داده‌ها مورد کاربرد قرار گرفته و به این واسطه امکانی فراهم گردیده تا خوشه‌ها در این چگالی همگن‌تر از سایر حالات باشند.

## ۴- نتایج و تحلیل نتایج

### • تحلیل چگالی خوشه‌ها

شکل ۷ روند تغییرات تعداد خوشه‌های تولیدی توسط الگوریتم خوشه‌بندی OPTICS را با افزایش حداقل نقاط همسایگی در  $\xi = 0.05$  نشان می‌دهد. الگوریتم OPTICS در  $MinPts = 4$ ، ۳۸ خوشه را شناسایی کرده و رفته رفته با افزایش حداقل نقاط همسایگی، تعداد خوشه‌ها به ۱۵ می‌رسد. این موضوع نشان می‌دهد که چگالی خوشه‌ها با تعداد آن‌ها رابطه عکس دارد؛ یعنی چنانچه حداقل نقاط همسایگی افزایش یابد، به طبع چگالی نیز در خوشه‌ای که شناسایی می‌شود افزایش یافته و به دنبال آن تعداد خوشه‌ها کاهش می‌یابد. در واقع با افزایش پارامتر حداقل نقاط همسایگی به بیش از ۴ نقطه، خوشه‌هایی که در  $MinPts = 4$  از حداقل نقاط همسایگی برخوردار بودند، دیگر به عنوان خوشه در نظر گرفته نشده و صرفاً به عنوان داده پرت شناخته می‌شوند. به علاوه، بر اساس این نمودار می‌توان بیان داشت که تعداد



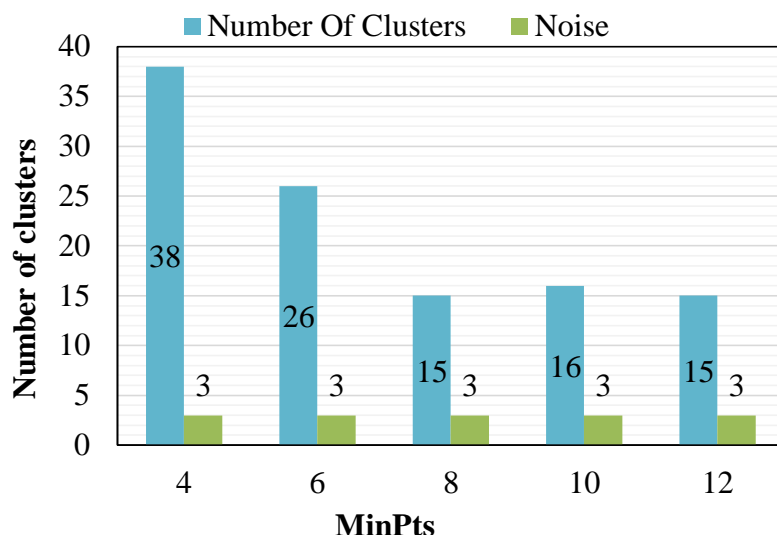


Fig. 7- Changes in the number of clusters and noises detected by OPTICS algorithm with different MinPts for  $\xi = 0.05$

شکل ۷- تغییرات تعداد خوشه‌ها و داده‌های پرت شناسایی شده توسط الگوریتم OPTICS با حداقل نقاط همسایگی مختلف برای  $\xi = 0.05$

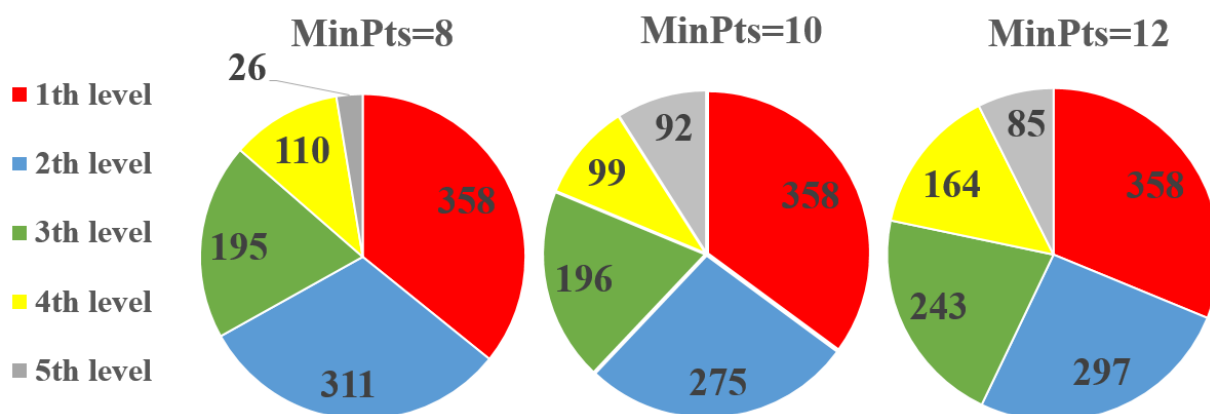


Fig. 8- Changes in the number of data at different levels of clustering by OPTICS algorithm at  $\xi = 0.05$  and at different MinPts

شکل ۸- تغییرات تعداد داده‌ها در سطوح مختلف خوشه‌بندی توسط الگوریتم OPTICS در  $\xi = 0.05$  و حداقل نقاط همسایگی مختلف

مناسبی برای اولویت‌بندی اجرای برنامه‌های بازسازی و نوسازی هستند. البته ذکر این نکته ضروری است که انتخاب هر سطح از سلسله مراتب خوشه‌ها برای بازسازی و نوسازی به عوامل و عناصر تصمیم‌گیرنده مختلف بستگی دارد؛ اما در حالت کلی خوشه‌ای که نرخ شکستگی بالایی دارد، به عنوان اولویت اصلی در طرح نوسازی و بازسازی مطرح می‌شود. همچنین، بعید به نظر می‌رسد که خوشه‌های خیلی بزرگ، نظر سرمایه‌گذار را برای اجرای طرح بازسازی و نوسازی جلب نماید.

به منظور تصمیم‌گیری بهتر جهت تعیین اولویت‌های بازسازی و نوسازی در یک طرح خوشه‌بندی مطلوب، می‌بایست تا حد ممکن، خوشه‌ها کوچک باشند. به طور خلاصه، وجود خوشه‌های با اندازه کوچک، امکان محاسبه نرخ شکست در خوشه‌ها و همگن بودن آن‌ها جزء شروطی است که برقراری هر یک بایستی در نتایج الگوریتم OPTICS مورد تحقیق قرار گیرد. علاوه بر این، احتمال می‌رود خوشه‌هایی که منطقه کوچک‌تری را پوشش می‌دهند، از منظر ویژگی بوله‌هایی که دچار حادثه شده‌اند، همگن‌تر باشند. بر همین مبنا می‌توان نتیجه گرفت که زیرخوشه‌های شناسایی شده در این روش، گزینه‌های

• تحلیل خوشه‌های حاصل از اجرای OPTICS با پارامترهای بهینه  $\xi = 0.05$  و  $MinPts = 10$

اهمیت قابل توجهی برخوردار خواهد بود. در سطح دوم الگوریتم OPTICS به صورت خودکار سطح اول را که خود مشتمل بر خوشه شماره ۱۶ بوده به چند خوشه کوچکتر تقسیم می‌کند. مطابق با شکل ۱۰-ب، مشاهده می‌شود که خوشه‌های شماره ۲، ۴، ۶ و ۱۳ به عنوان زیرخوشه‌های این سطح شناسایی شده‌اند. به علاوه، از آنجایی که با کوچکتر شدن خوشه‌ها، نرخ شکست افزایش می‌یابد و به دنبال آن امکان کشف مناطق بحرانی آسان‌تر می‌گردد، از این رو فرآیند سطح‌بندی خوشه‌ها توسط الگوریتم OPTICS تا سطوح بعدی ادامه یافته و سطح سوم نیز به صورت خودکار شناسایی خواهد شد.

بر مبنای اصول حاکم بر الگوریتم OPTICS و همچنین منطبق بر نمودار فاصله دسترسی‌پذیری (شکل ۹)، تمامی خوشه‌های سطح دوم نیز به خوشه‌های کوچکتری تقسیم گردیدند. در واقع خوشه شماره ۱ به عنوان زیرمجموعه خوشه شماره ۲ شناخته شده و افزون بر این، خوشه‌های ۳ و ۵ در داخل خوشه شماره ۴ و خوشه‌های شماره ۷ و ۱۳ نیز درون خوشه شماره ۱۲ شناسایی شده‌اند که چگونگی توزیع هر یک از آن‌ها در شکل ۱۱-الف نمایان گردیده است. همچنین، لازم به ذکر است که نرخ شکست و تراکم حوادث شکستگی رخ داده در مناطق این خوشه‌ها طبیعتاً بالاتر از نواحی متناظر با خوشه‌های سطح دوم می‌باشد. در سطح چهارم نیز خوشه شماره ۱۳ که پیشتر به عنوان یکی از زیرخوشه‌های سطح سوم شناسایی شده بود، به دو خوشه شماره ۹ و ۱۴ تقسیم می‌شود.

الگوریتم OPTICS پس از محاسبه مقادیر فاصله دسترسی‌پذیری هر یک از داده‌ها، آن‌ها را در نموداری مرتب نموده که تحت عنوان نمودار فاصله دسترسی‌پذیری عنوان می‌شود. همچنین، می‌توان مشاهده کرد که چگونگی نوسانات موجود در این نمودار با تأثیرگذاری قابل توجهی در مبانی خوشه‌بندی و سطح‌بندی خوشه‌ها همراه است. با توجه به تعیین مقادیر بهینه پارامترهای تنظیمی معادل  $\xi = 0.05$  و  $MinPts = 10$  برای الگوریتم OPTICS در این پژوهش، روند تغییرات فاصله دسترسی‌پذیری در شکل ۹ قابل مشاهده می‌باشد. در این نمودار، مناطق مقعر نشانگر خوشه‌هایی است که توسط الگوریتم شناسایی شده و بررسی دقیق این مناطق بر وجود زیرخوشه‌ها یا به عبارتی شناسایی خوشه‌های تودرتو صحنه می‌گذارد.

همانطور که پیشتر در شکل ۸ نشان داده شد، الگوریتم OPTICS به ازای مقادیر  $\xi = 0.05$  و  $MinPts = 10$  خوشه‌های شناخته شده را در ۵ سطح طبقه‌بندی می‌نماید. شکل ۱۰-الف، سطح اول خوشه‌بندی حاصل از الگوریتم OPTICS را نشان می‌دهد. در این سطح، الگوریتم تنها داده‌های پرت را شناسایی می‌کند و مابقی داده‌های حوادث را در داخل یک خوشه قرار می‌دهد. این خوشه تقریباً تمام شبکه را در بر دارد، بنابراین برای بازسازی و نوسازی گزینه مناسبی نمی‌تواند باشد. بر همین مبنا، تحلیل دیگر سطوح خوشه‌بندی (سطح دوم به بعد) از

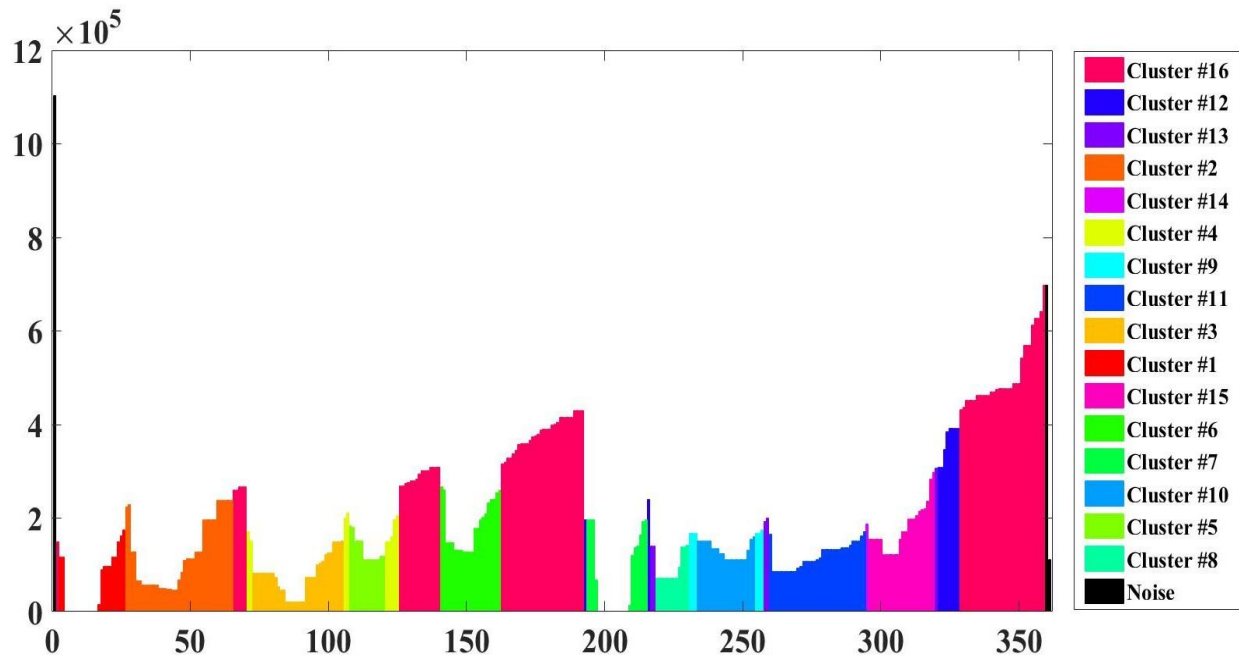


Fig. 9- The reachability-plot for the implementation of the OPTICS algorithm for  $MinPts = 10$  &  $\xi = 0.05$   
 شکل ۹- منحنی فاصله دسترسی‌پذیری حاصل از اجرای الگوریتم OPTICS برای  $MinPts = 10$  و  $\xi = 0.05$

خوشه شماره ۱۴ جای می‌گیرند. مجموعاً، تمامی موارد یاد شده در شکل ۱۲ قابل مشاهده می‌باشد.

در ادامه، تعیین اولویت‌های نوسازی و بازسازی شبکه مورد بررسی، با استفاده از تحلیل نرخ شکستگی و تراکم داده‌های موجود در خوشه‌ها و همچنین ارزیابی اثرگذاری پارامترهای مرتبط با فیزیک شبکه در نرخ حوادث صورت پذیرفته است. برای این منظور، پس از محاسبه نرخ شکست در هر خوشه بر مبنای تعداد حوادث شکستگی رخ داده در سال در هر ۱۰۰ کیلومتر لوله موجود در آن خوشه، تراکم داده‌های موجود در هر خوشه نیز محاسبه گردیده است (شکل ۱۳).

مجموعه این زیرخوشه‌ها نیز سطح چهارم را به وجود می‌آورند که چگونگی توزیع آن‌ها در محدوده شبکه توزیع منطقه چهار در شکل ۱۱-ب قابل مشاهده است.

در سطح پنجم که به عنوان آخرین سطح خوشه‌بندی تلقی می‌شود، هر کدام از دو خوشه شماره ۹ و ۱۴ (خوشه‌های شناسایی شده در سطح پیشین) خود به دو خوشه مجزای دیگر تقسیم می‌گردند. به عبارتی در این حالت، فرآیند خوشه‌بندی به گونه‌ای توسعه یافته که خوشه‌های شماره ۸ و ۱۰ درون خوشه شماره ۹ و خوشه‌های ۱۱ و ۱۵ نیز داخل

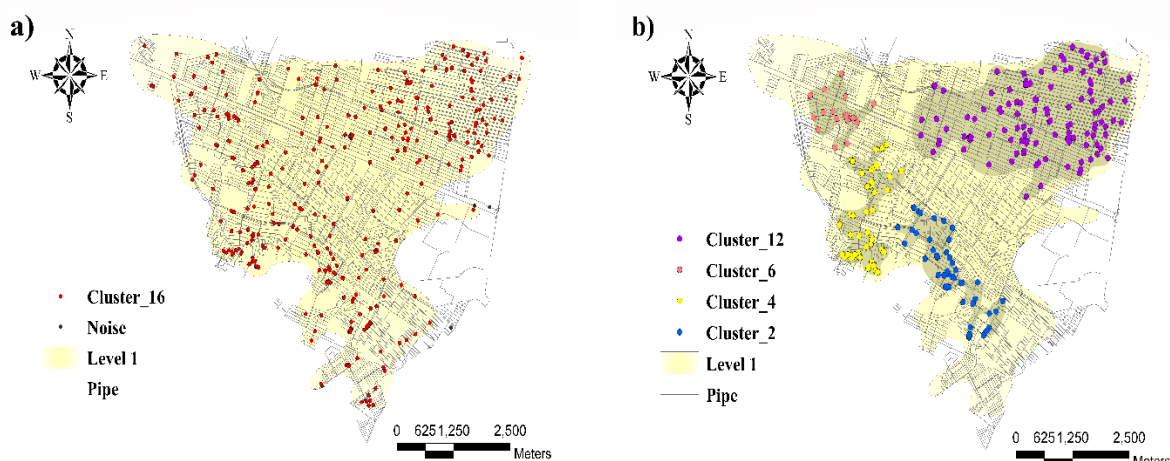


Fig. 10- Levels identified by the OPTICS algorithm for  $MinPts = 10$  &  $\xi = 0.05$ , a) First level, b) Second level  
 شکل ۱۰- نمایش گرافیکی سطوح خوشه‌بندی توسط الگوریتم OPTICS به ازای  $\xi = 0.05$  و  $MinPts = 10$  (الف) سطح اول، (ب) سطح دوم

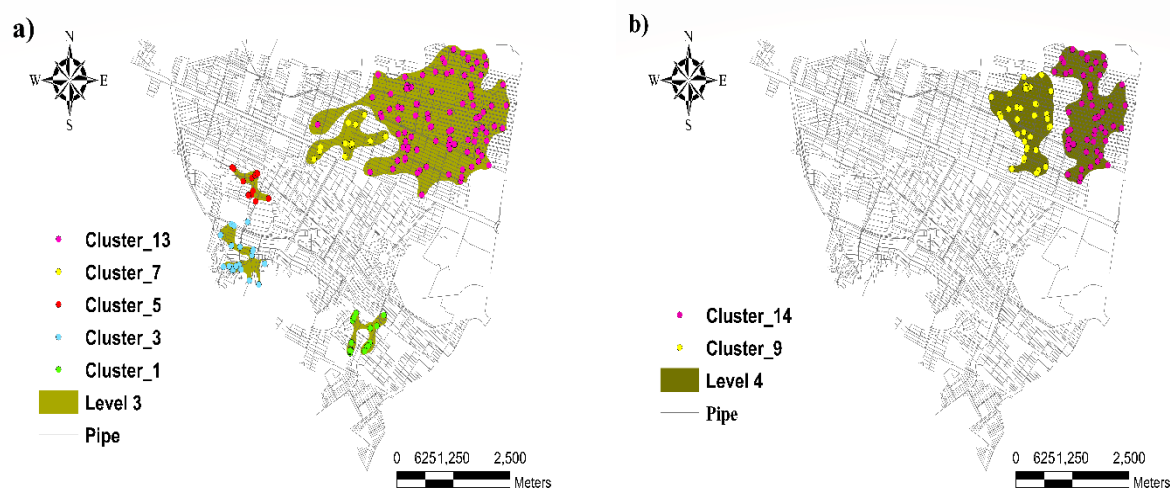


Fig. 11- Levels identified by the OPTICS algorithm for  $MinPts = 10$  &  $\xi = 0.05$ , a) 3<sup>th</sup> level, b) 4<sup>th</sup> level  
 شکل ۱۱- نمایش گرافیکی سطوح خوشه‌بندی توسط الگوریتم OPTICS به ازای  $\xi = 0.05$  و  $MinPts = 10$  (الف) سطح سوم، (ب) سطح چهارم

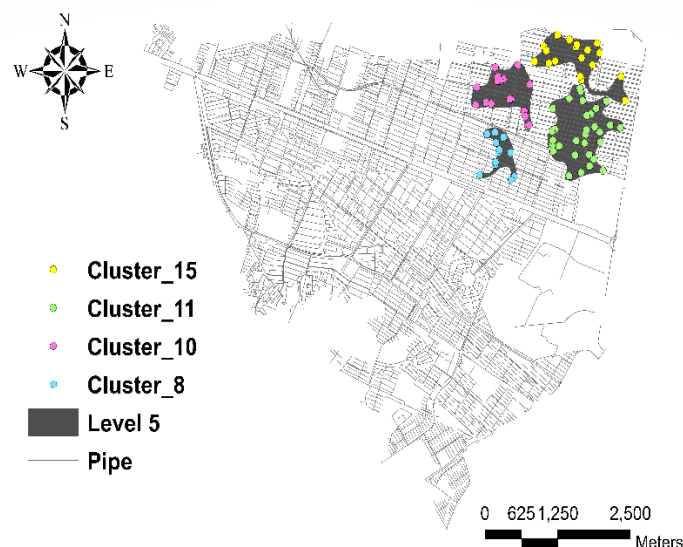


Fig. 12- 5<sup>th</sup> level identified by the OPTICS algorithm, MinPts = 10 &  $\xi = 0.05$

شکل ۱۲- نمایش گرافیکی سطح پنجم خوشه‌بندی شناسایی شده توسط الگوریتم OPTICS،  $\xi = 0.05$  و MinPts = 10

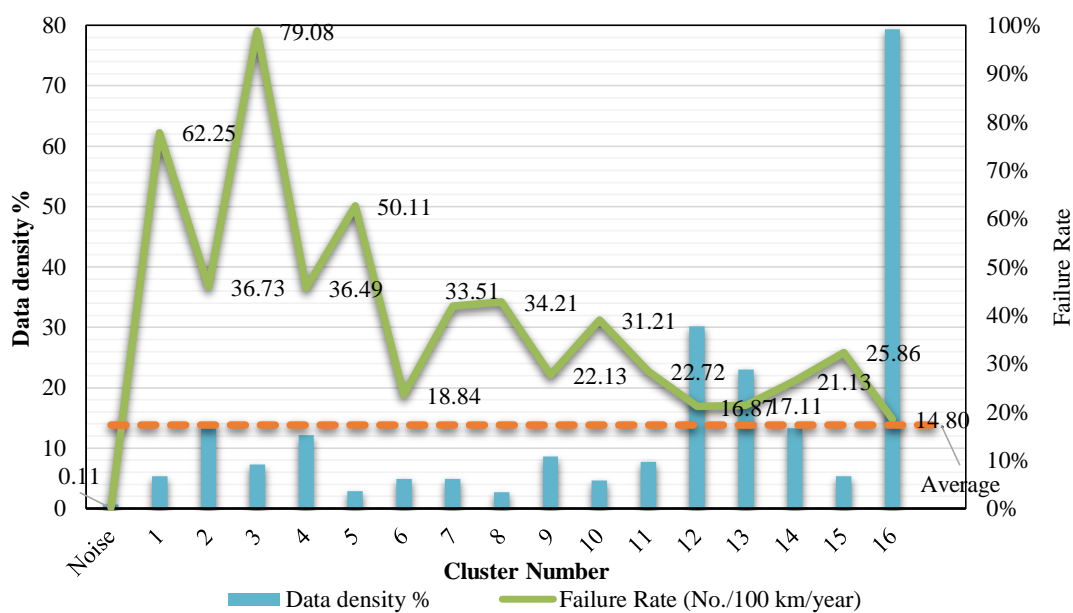


Fig. 13- Failure rate plot and data density in each cluster based on the results of OPTICS algorithm, MinPts = 10 &  $\xi = 0.05$

شکل ۱۳- نمودار نرخ شکستگی و تراکم داده‌های موجود در هر خوشه بر اساس نتایج الگوریتم OPTICS،  $\xi = 0.05$  و MinPts = 10

اختصاص داده‌اند، جزء خوشه‌هایی هستند که کمترین نرخ شکستگی را دارا می‌باشند. از طرفی دیگر، خوشه شماره ۵ یا به عبارتی دیگر زیرخوشه مرتبط با خوشه شماره ۴ نیز در رتبه سوم از جهت میزان نرخ شکست قرار گرفته و در مجموع، نرخ شکستگی در خوشه ۴ دارای مقداری بیشتر از نرخ شکست میانگین است. از طرف دیگر، بر مبنای این نمودار مشخص می‌گردد که تراکم داده‌ها در خوشه شماره ۵ کمتر از ۵ درصد و در خوشه شماره ۳ نیز کمتر از ۱۵ درصد می‌باشد. این در

مطابق با این نمودار، بیشترین نرخ شکست متعلق به خوشه شماره ۳ می‌باشد که در سطح سوم جای داشته و یکی از زیرمجموعه‌های (زیرخوشه) خوشه شماره ۴ به شمار می‌رود. همچنین مشاهده می‌شود که خوشه شماره ۱ که در محدوده جغرافیایی خوشه ۲ جای دارد، در حالی که ۱۰ درصد از داده‌ها را به خود اختصاص داده است در رتبه دوم نرخ شکستگی قرار گرفته است. این در حالی است که خوشه‌های شماره ۱۲ و ۱۳ که هر دو بیش از ۳۰ درصد از تراکم داده‌ها را به خود

اساس شکل ۱۵، اکثر حوادث این خوشه بر روی لوله‌های آزرست سیمان و پلی‌اتیلن رخ داده که قطری کمتر از ۱۰۰ میلی‌متر دارند. بنابراین اولویت اصلی برنامه بازسازی و نوسازی در شبکه آب منطقه مورد مطالعه بر اساس آنچه که الگوریتم OPTICS نشان می‌دهد، مجموعه خوشه‌های ۳، ۱ و ۵ می‌باشد که تقریباً در هر سه خوشه، قطر کمتر از ۱۰۰ میلی‌متر و لوله‌هایی که از جنس آزرست سیمان و پلی‌اتیلن هستند، به عنوان عوامل اصلی بروز نرخ بالای شکست در این مناطق عمل نموده‌اند. همچنین با انجام تحلیل‌های مشابه، روشن گردید که مجموعه خوشه‌های شماره ۷، ۸ و ۱۰ به عنوان رده دوم اولویت‌های بازسازی و نوسازی قابل طرح می‌باشند. لازم به ذکر است که منظور از رده اول اولویت‌های بازسازی و نوسازی، اولویت‌های اصلی جهت اجرای عملیات بازسازی شبکه بوده و همچنین هدف از تعیین خوشه‌های رده دوم، تفکیک مناطقی است که در قیاس با خوشه‌های رده اول از درجه اهمیت پایین‌تری جهت پیاده‌سازی برنامه‌های نوسازی برخوردارند.

#### ۵- جمع‌بندی و نتیجه‌گیری

اولویت‌بندی حوادث شکستگی در شبکه توزیع آب برای اجرای برنامه‌های بازسازی و نوسازی امری ضروری است. در این مطالعه، از الگوریتم خوشه‌بندی مبتنی بر چگالی OPTICS جهت شناسایی پهنه‌های اولویت‌دار جهت اجرای برنامه‌های بازسازی و نوسازی در یکی از پهنه‌های بهره‌بردار شبکه توزیع آب شهر مشهد بهره‌برده شد.

حالی است که همچنان خوشه‌های زیادی وجود دارند که علی‌رغم دارا بودن تراکم بیشتر، نرخ شکست کمتری را ثبت نموده‌اند.

از سوی دیگر، نحوه اثرگذاری پارامترهای مرتبط با فیزیک شبکه (از جمله مؤلفه‌های قطر و جنس) بر روی متوسط نرخ شکستگی خوشه‌ها نیز به ترتیب در نمودارهای ۱۴ و ۱۵ قابل بررسی است. بر اساس شکل ۱۴، خوشه ۳ تنها از حوادثی تشکیل شده که لوله‌های محل وقوع حادثه همگی قطری کمتر از ۱۰۰ میلی‌متر دارند و از این نظر، خوشه شماره ۳ خوشه‌ای کاملاً همگن ارزیابی می‌شود. با این حال، منطق بر شکل ۱۵ و از منظر جنس لوله‌ها، می‌توان مشاهده کرد که حوادث این خوشه در مجموعه‌ای از لوله‌های فولادی، آزرست سیمانی و پلی‌اتیلنی رخ داده است. بنابراین با توجه به تنوع جنس، می‌توان بیان داشت که پارامتر قطر، یکی از عوامل تأثیرگذار در افزایش نرخ شکستگی و افزایش حوادث در منطقه جغرافیایی متناظر با خوشه شماره ۳ است. افزون بر این، اکثر حوادث در خوشه شماره ۵ نیز بر روی لوله‌های با اقطار کمتر از ۱۰۰ میلی‌متر رخ داده که از جنس پلی‌اتیلن و آزرست سیمان ساخته شده‌اند و بر همین مبنا می‌توان پارامتر اثرگذار در خوشه شماره ۵ را نیز تعیین نمود. با بررسی سایر خوشه‌ها نیز مشخص گردید که خوشه شماره ۱ که خود زیرمجموعه خوشه شماره ۲ می‌باشد، در رتبه دوم از لحاظ میزان نرخ شکست قرار گرفته است. خوشه شماره ۲ علی‌رغم در برداشتن تراکم بالای حوادث، نرخ شکست پایین‌تری را در قیاس با زیرخوشه‌های خود دارا است. خوشه شماره ۱ نیز تنها ۱۰ درصد از حوادث را به خود اختصاص داده و بر

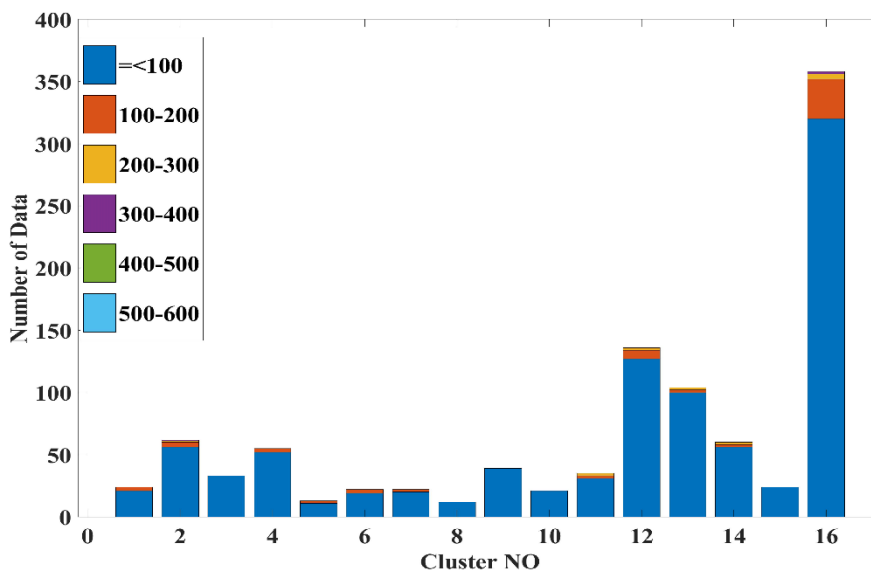


Fig. 14- Breakdown of pipes in each cluster for different pipe diameter,  $\text{MinPts} = 10$  &  $\xi = 0.05$

شکل ۱۴- تفکیک لوله‌های موجود در هر خوشه برای قطرهای مختلف لوله،  $\xi = 0.05$  و  $\text{MinPts} = 10$

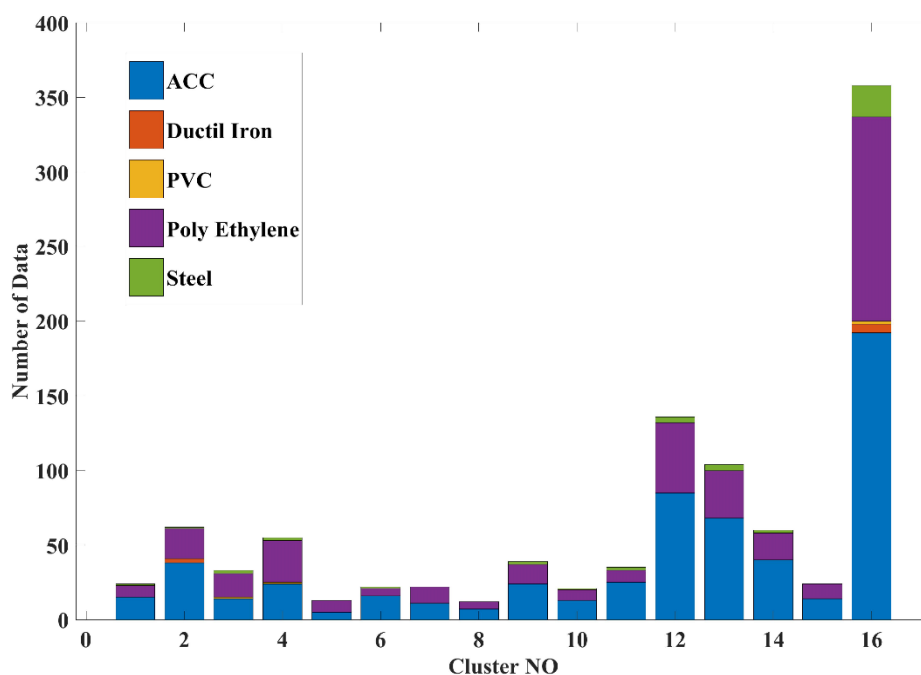


Fig. 15- Breakdown of pipes in each cluster for different pipe material, MinPts = 10 &  $\xi = 0.05$   
 شکل ۱۵- تفکیک لوله‌های موجود در هر خوشه برحسب جنس لوله‌ها،  $\xi = 0.05$  و MinPts = 10

داده‌ها از منظر ویژگی‌هایی چون فشار، بارهای محیطی، سن لوله‌ها و غیره می‌توان به علل اصلی حوادث در شبکه نیز پی برد. همچنین، از آنجایی که در این پژوهش تنها از دو بعد طول و عرض جغرافیایی بهره برده شده است، پیشنهاد می‌شود که با افزایش بعد داده‌های حوادث و گنجانیدن ویژگی‌های مذکور به بررسی بیشتر تأثیرات عوامل زمینه‌ای پرداخته شود. همچنین، خوشه‌های شناسایی شده نیز می‌توانند از منظر اقتصادی نیز تحلیل گردند تا اولویت‌های مهم‌تر برای سرمایه‌گذاری انتخاب گردند. از آنجایی که OPTICS حالت تعمیم یافته‌ای از الگوریتم DBSCAN می‌باشد، پیشنهاد می‌گردد که این رویکرد با الگوریتم DBSCAN با مجموعه داده یکسان بار دیگر تکرار شود تا مقایسه جامعی میان این دو الگوریتم برای بهره‌برداران در شبکه‌های توزیع آب به عمل آید. در این مقاله سعی شده است پارامترهای ورودی به الگوریتم بر اساس اصول مشخصی انجام شود تا سلیقه بهره‌بردار مربوطه بر نتایج پایانی دخیل نگردد، اما پیشنهاد می‌گردد تا به کمک روش‌های دیگری از جمله الگوریتم‌های بهینه‌سازی، نیز این موضوع بررسی شود. با توجه به توضیحات ارائه شده در این پژوهش می‌توان اذعان داشت که الگوریتم OPTICS چه در تشخیص داده‌های پرت و چه در یافتن اولویت‌های بازسازی و نوسازی شبکه توزیع آب، خصوصاً از حیث تعیین خودکار اولویت‌ها بر اساس سطح‌بندی الگوریتم و تطابق این موضوع با بالا بودن نرخ حوادث شکستگی در این خوشه‌ها، توانایی قابل قبولی دارد.

با توجه به نتایج بدست آمده برای الگوریتم OPTICS و در نظر گرفتن شاخص‌های کارآمدی آن، حداقل نقاط همسایگی MinPts = 10 و حداقل شیب مناسب جهت تفکیک خوشه‌ها در نمودار فاصله دسترسی‌پذیری  $\xi = 0.05$  بدست آمد. الگوریتم OPTICS برای مجموعه داده‌های حاضر ۱۶ خوشه را شناسایی کرد. این تعداد خوشه در ۵ سطح قرار دارند که با بالاتر رفتن هر سطح، خوشه‌ها به خوشه‌های کوچک‌تر که نرخ حوادث شکستگی بیشتری دارند تقسیم می‌شوند. بر اساس مقایسه نرخ شکست در خوشه‌های حاصل از الگوریتم OPTICS با نرخ میانگین حوادث شکستگی در کل شبکه، ۳ خوشه با نرخ شکست ۰٫۷۹/۱، ۶۲/۲ و ۵۱/۱ حادثه در هر ۱۰۰ کیلومتر طول لوله در یک سال، اولویت‌های اصلی برنامه بازسازی و نوسازی باید باشند. همچنین سه خوشه دیگر نیز با نرخ شکست ۳۳/۵، ۳۴/۲ و ۳۱/۲ در اولویت دوم برنامه بازسازی و نوسازی قرار می‌گیرند. این در حالی است که نرخ شکست میانگین شبکه برابر با ۱۴/۸ حادثه در ۱۰۰ کیلومتر در سال بود. طبق تحلیل قطر و جنس صورت گرفته بر روی خوشه‌های حاصل از این الگوریتم نیز نشان داد که برای تدوین برنامه بازسازی در این خوشه‌ها باید توجه بیشتر بر لوله‌هایی متمرکز شود که قطری کمتر از ۱۰۰ میلی‌متر دارند و از جنس پلی‌اتیلن و آزیست سیمانی هستند. تحلیل قطر و جنس صورت گرفته نیز نشان داد که الگوریتم OPTICS در شناسایی خوشه‌های همگن و متمایز توانایی بالایی دارد. این امر نشان داد که با کامل‌تر بودن مجموعه

## پی نوشت‌ها

- 1- Water Distribution Network
- 2- Exploratory Analysis
- 3- Failure Models
- 4- Asset Management Models
- 5- Decision-Making Models
- 6- Density-Based Spatial Clustering of Applications with Noise
- 7- Expectation Maximization
- 8- Spatial Autocorrelation Analysis

## ۶- مراجع

- Abokifa AA and Sela L (2019) Identification of spatial patterns in water distribution pipe failure data using spatial autocorrelation analysis. *Journal of Water Resources Planning and Management* 145(12)
- Ankerst M, Breunig MM, Kriegel H-P, and Sander J (1999) OPTICS: Ordering points to identify the clustering structure. *ACM Sigmod Record* 28(2):49–60
- Aryayee A, Moslehi I, Jalili-Ghazizadeh M, Komeily M, Hoda-Atashkar M, and Maghrebi H (2020) Density based clustering analysis of water main break events to determine prior area for rehabilitation. 3rd Iran Water and Wastewater Science and Engineering Congress, 24-26 November, 2020, Shiraz University, Shiraz, Iran (In Persian)
- Baird GM (2010) A game plan for aging water infrastructure. *American Water Works Association* 102(4):74–82
- De Oliveira DP, Garrett JH, and Soibelman L (2011) A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Advanced Engineering Informatics* 25(2):380–389
- Engelhardt M, Savic D, Skipworth P, Cashman A, Saul A, and Walters G (2003) Whole life costing: Application to water distribution network. *Water Science and Technology: Water Supply* 3(1–2):87–93
- Ester M, Kriegel H-P, Sander J, and Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 226–231
- Kabir G, Tesfamariam S, and Sadiq R (2015) Bayesian model averaging for the prediction of water main failure for small to large Canadian municipalities. *Canadian Journal of Civil Engineering, Canadian Science Publishing* 43(3):233–240
- Kleiner Y and Rajani B (2001) Comprehensive review of structural deterioration of water mains: Statistical models. *Urban Water* 3(3):131–150
- Kleiner Y and Rajani B (2010) I-WARP: Individual water main renewal planner. *Drinking Water Engineering and Science* 3(1):71–77
- Moslehi I, Jalili-Ghazizadeh M, and Yousefi Khoshqalb E (2019) Flow data monitoring in water distribution networks using density-based clustering methods. *Iran-Water Resources Research* 15(3):62–77 (In Persian)
- Nazemi A and Madani K (2018) Urban water security: Emerging discussion and remaining challenges. *Sustainable Cities and Society, Elsevier B.V.* 41(17):925–928
- Oliveira D, Garrett JH, and Soibelman L (2009) Spatial clustering analysis of water main break events. *Proceedings of the 2009 ASCE International Workshop on Computing in Civil Engineering*, 338–347
- Salehi S, Jalili Ghazizadeh M, and Tabesh M (2018) A comprehensive criteria-based multi-attribute decision-making model for rehabilitation of water distribution systems. *Structure and Infrastructure Engineering, Taylor & Francis* 14(6):743–765
- Sander J, Ester M, Kriegel H-P, and Xu X (1998) Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery* 2(2):169–194
- Sattar AMA, Ertuğrul ÖF, Gharabaghi B, McBean EA, and Cao J (2019) Extreme learning machine model for water network management. *Neural Computing and Applications* 31(1):157–169
- Scheidegger A, Leitão JP, and Scholten L (2015) Statistical failure models for water distribution pipes- A review from a unified perspective. *Water Research* 83:237–247
- Xu Q, Chen Q, Li W, and Ma J (2011) Pipe break prediction based on evolutionary data-driven methods with brief recorded data. *Reliability Engineering and System Safety, Elsevier* 96(8):942–948