ARVIND SAI DOODA A20553046

**READING ASSIGNMENT -3 CLUSTERING**

**Reference Paper :**

**Denpeak: RODRIGUEZ and LAIO, Clustering by fast search and find of density peaks, Science, 2014.**





intrinsic and extrinsic contributions depends on the sample quality (such as the doping density and the amount of disorder). Studies of the dependence on temperature and on disorder are therefore required to better understand the doping density dependence of the VHE. Furthermore, a more accurate determination of $\sigma_H$ that takes into account the fringe fields in our Hall bar device may be needed for a better quantitative comparison.

**REFERENCES AND NOTES**

1. A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Rev. Mod. Phys.* **81**, 109–162 (2009).
2. A. Rycerz, J. Tworzydło, C. W. J. Beenakker, *Nat. Phys.* **3**, 172–175 (2007).
3. A. R. Akhmerov, C. W. J. Beenakker, *Phys. Rev. Lett.* **98**, 157003 (2007).
4. D. Xiao, W. Yao, Q. Niu, *Phys. Rev. Lett.* **99**, 236809 (2007).
5. W. Yao, D. Xiao, Q. Niu, *Phys. Rev. B* **77**, 235406 (2008).
6. D. Xiao, G.-B. Liu, W. Feng, X. Xu, W. Yao, *Phys. Rev. Lett.* **108**, 196802 (2012).
7. Y. J. Zhang, T. Oka, R. Suzuki, J. T. Ye, Y. Iwasa, *Science* **344**, 725–728 (2014).
8. K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, *Phys. Rev. Lett.* **105**, 136805 (2010).
9. A. Splendiani *et al.*, *Nano Lett.* **10**, 1271–1275 (2010).
10. T. Cao *et al.*, *Nat. Commun.* **3**, 887 (2012).
11. K. F. Mak, K. He, J. Shan, T. F. Heinz, *Nat. Nanotechnol.* **7**, 494–498 (2012).
12. G. Sallen *et al.*, *Phys. Rev. B* **86**, 081301 (2012).
13. S. Wu *et al.*, *Nat. Phys.* **9**, 149–153 (2013).
14. H. Zeng, J. Dai, W. Yao, D. Xiao, X. Cui, *Nat. Nanotechnol.* **7**, 490–493 (2012).
15. D. Xiao, M.-C. Chang, Q. Niu, *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
16. X. Li, F. Zhang, Q. Niu, *Phys. Rev. Lett.* **110**, 066803 (2013).
17. S. Murakami, N. Nagaosa, S.-C. Zhang, *Science* **301**, 1348–1351 (2003).
18. J. Sinova *et al.*, *Phys. Rev. Lett.* **92**, 126603 (2004).
19. Y. K. Kato, R. C. Myers, A. C. Gossard, D. D. Awschalom, *Science* **306**, 1910–1913 (2004).
20. J. Wunderlich, B. Kaestner, J. Sinova, T. Jungwirth, *Phys. Rev. Lett.* **94**, 047204 (2005).
21. J. Wunderlich *et al.*, *Science* **330**, 1801–1804 (2010).
22. N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, N. P. Ong, *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
23. Materials and methods are available as supplementary materials on Science Online.
24. The skew scattering contribution, which is important only for high-mobility devices (*22*), is neglected in $MoS_2$ devices with relatively low mobility.
25. T. Cheiwchanchamnangij, W. R. L. Lambrecht, *Phys. Rev. B* **85**, 205302 (2012).
26. B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, A. Kis, *Nat. Nanotechnol.* **6**, 147–150 (2011).
27. L. J. van der Pauw, *Philips Techn. Rev.* **20**, 220–224 (1958).

optical data on monolayer $MoS_2$. This research was supported by the Kavli Institute at Cornell for Nanoscale Science and the Cornell Center for Materials Research [National Science Foundation (NSF) DMR-1120296]. Additional funding was provided by the Air Force Office of Scientific Research (FA9550-10-1-0410) and the Nano-Material Technology Development Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (2012M3A7B4049887). Device fabrication was performed at the Cornell NanoScale Science and Technology Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by NSF (grant ECCS-0335765). K.L.M. acknowledges support from the NSF Integrative Graduate Education and Research Traineeship program (DGE-0654193) and the NSF Graduate Research Fellowship Program (DGE-1144153).

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/344/6191/1489/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*34*–*37*)

24 December 2013; accepted 23 May 2014
10.1126/science.1250140

MACHINE LEARNING

# Clustering by fast search and find of density peaks

**Alex Rodriguez and Alessandro Laio**

**Cluster analysis is aimed at classifying elements into categories on the basis of their similarity. Its applications range from astronomy to bioinformatics, bibliometrics, and pattern recognition. We propose an approach based on the idea that cluster centers are characterized by a higher density than their neighbors and by a relatively large distance from points with higher densities. This idea forms the basis of a clustering procedure in which the number of clusters arises intuitively, outliers are automatically spotted and excluded from the analysis, and clusters are recognized regardless of their shape and of the dimensionality of the space in which they are embedded. We demonstrate the power of the algorithm on several test cases.**

Clustering algorithms attempt to classify elements into categories, or clusters, on the basis of their similarity. Several different clustering strategies have been proposed (*1*), but no consensus has been reached even on the definition of a cluster. In K-means (*2*) and K-medoids (*3*) methods, clusters are groups of data characterized by a small distance to the cluster center. An objective function, typically the sum of the distance to a set of putative cluster centers, is optimized (*3*–*6*) until the best cluster centers candidates are found. However, because a data point is always assigned to the nearest center, these approaches are not able to detect nonspherical clusters (*7*). In distribution-based algorithms, one attempts to reproduce the observed realization of data points as a mix of predefined tion. This method allows the finding of nonspherical clusters but works only for data defined by a set of coordinates and is computationally costly.

Here, we propose an alternative approach. Similar to the K-medoids method, it has its basis only in the distance between data points. Like DBSCAN and the mean-shift method, it is able to detect nonspherical clusters and to automatically find the correct number of clusters. The cluster centers are defined, as in the mean-shift method, as local maxima in the density of data points. However, unlike the mean-shift method, our procedure does not require embedding the data in a vector space and maximizing explicitly the density field for each data point.

The algorithm has its basis in the assumptions that cluster centers are surrounded by neighbors

**SUMMARY :**

- Issue: The challenge lies in effectively clustering data points when dealing with non-spherical clusters, uncertain cluster numbers, and outlier detection.

- Solution: The authors introduce a density-based clustering method, where cluster centers are identified based on local density and distance from denser points. Each data point is assigned to the cluster of its nearest, denser neighbor.

- Outcomes: This approach offers intuitive cluster number determination, the ability to handle various cluster shapes, and outlier detection. It's computationally efficient and resilient to noise. Experimental results on both synthetic and real datasets validate its efficacy.

**PROS :**

- Intuitive cluster center identification through density peaks.
- Automatic cluster number determination.
- Effective handling of non-spherical clusters and outliers.
- Computational efficiency with a low misclassification rate.
- Strong performance in high-dimensional data scenarios.

**CONS:**

- Inaccurate density estimation for small datasets.
- Trial-and-error process for selecting the cutoff distance (dc).
- Dependency on a well-defined distance metric between data points.
- Difficulty in identifying clusters when the number of clusters is close to the dataset size.
- High computational cost for large datasets due to the need for a full distance matrix.