NETWORK CYCLE FEATURES: APPLICATION TO COMPUTER-AIDED GLEASON GRADING OF PROSTATE CANCER HISTOPATHOLOGICAL IMAGES

Parmeshwar Khurd, Leo Grady and Ali Kamen

Siemens Corporate Research, Princeton, NJ, USA

Summer Gibbs-Strauss, Elizabeth M. Genega, John V. Frangioni *

Beth Israel Deaconess Medical Center, Boston, MA, USA

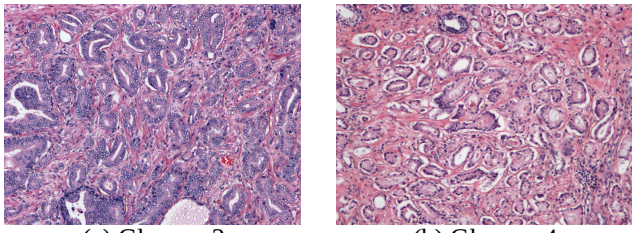
ABSTRACT

Features extracted from cell networks have become popular tools in histological image analysis. However, existing features do not take sufficient advantage of the cycle structure present within the cell networks. We introduce a new class of network cycle features that take advantage of such structures. We demonstrate the utility of these features for automated prostate cancer scoring using histological images. Prostate cancer is commonly scored by pathologists using the Gleason grading system and our automated system based upon network cycle features serves an important need in making this process less labor-intensive and more reproducible. Our system first extracts the cells from the histological images, computes networks from the cell locations and then computes features based upon statistics for the different cycles present in these networks. Using an SVM (Support Vector Machine) classifier on these features, we demonstrate the efficacy of our system in distinguishing between grade 3 and grade 4 prostate tumors. We also show the superiority of our approach over previously developed systems for this problem based upon texture features, fractal features and alternative network features.

***Index Terms*—** Gleason grading, prostate cancer, network features, classification

1. INTRODUCTION

Prostate cancer is the most common cancer, excluding skin cancer, and the second leading cause of death in American men, after lung cancer. Candidates suspected to have prostate cancer commonly undergo tissue biopsy in order to assess the presence and aggressiveness of cancer. The biopsied tissue samples are imaged with a microscope after hematoxylin and eosin (H&E) staining and assigned tumor grades according to the Gleason grading system (grades 1-5). In Fig. 1, we have shown H&E images of prostate cancer tissue samples corresponding to Gleason grades 3 and 4. The Gleason grade characterizes tumor differentiation, i.e. the degree to which the tumor resembles healthy tissue. The sum of the primary and the secondary Gleason grades yields the Gleason score, the single most important prognostic indicator for prostate cancer patients. The Gleason score plays an important role in deciding the future course of treatment.



(a) Gleason 3 (b) Gleason 4

Fig. 1. H&E images of cancerous prostate tissue

However, the assignment of Gleason scores is a time-consuming, error-prone process that depends upon the samples obtained during core biopsy as well as on the expertise of the pathologist. One way to validate the Gleason score obtained during core biopsy is to re-calculate this score in patients who undergo radical prostatectomy, thereby eliminating the sampling error. This form of validation is also required in order to study the correlation of prostate cancer biomarkers observed in other macroscopic imaging modalities with the Gleason score [1]. Computer-aided Gleason grading becomes essential when we need to assign tumor grades to the entire prostate specimen. Therefore, we have developed a computer-aided system to assign Gleason grades in an automatic and reproducible manner.

Our system is based upon a novel class of network cycle features extracted from the cell networks present in H&E images of prostate cancer. Our novel class of network features consists of statistics computed on the different cycles present within Delaunay triangulation of the cell centroid locations as well as other cell networks. The utility of features based upon cyclical structures is evident from the fact that cells in human tissue naturally form different structures (e.g., glands, ducts, etc.), each associated with its own characteristic cycle structure. Moreover, prostate cancer [2], as well as other forms of cancer [3], are characterized by the existence

*This work was supported by grant 1R01CA134493-01A1 from the NIH.

of modified structures. For instance, grade 3A prostate cancer [2] shows the existence of glands with elongated, angular and twisted forms and grade 3C prostate cancer exhibits papillary and cribriform epithelium, whereas grade 4 prostate cancer exhibits fused glands, creating masses, cords, or chains. Therefore, we can reasonably expect the network cycle features to capture the discriminatory patterns existing in grade 3 vs. grade 4 cancer.

In Sec. 2, we have provided a detailed description of our network cycle features and of our entire system. In Sec. 3, we have demonstrated the efficacy of our system in distinguishing between Gleason grades 3 and 4, and provided comparisons with alternative methods. We conclude with some final remarks in Sec. 4.

2. METHODS

2.1. Prostate Cancer Grading System Overview

Our system consists of: (1) cell detection, (2) cell network computation via Delaunay triangulation and nearest-neighbor techniques, (3) cycle extraction and cycle statistics computation and computation of other network-based features, and (4) SVM classification [4] using the network features.

For cell detection, we use a linear SVM operating in a sliding window manner, along with a non-max suppression heuristic. We use a few manual annotations of nuclei and non-nuclei to train this linear SVM. The advantage of using a linear SVM is that the SVM scores can be rapidly computed with convolutions. Since we use all three channels (R, G, B), this amounts to three image convolutions. We then rank the positive (nuclei) SVM outputs by their probability scores and sequentially pick the highest ranking output. To impose non-max suppression, after a nuclei is picked, the probability scores in a tolerance circle around it are set to zero. Examples of our cell detection results are shown in Fig. 2.

Example Delaunay networks computed from the cell detection results are shown in Fig. 2. We shall describe the cycle extraction and statistics computation in greater detail in Sec. 2.3. Given the various network features from Sec. 2.3, we use an SVM with a radial-basis-function kernel in order to distinguish between the different Gleason grades.

2.2. Related Work

Pioneering work on the use of topological cell network features for cancer grading in histological imaging was done in [5], although they did not consider the Gleason grading problem. We describe some of their network features in Sec. 2.3 and compare them with our novel network features in Sec. 3. Automated Gleason grading was the focus in [6, 7, 8, 9, 10]. The work in [8] used global fractal dimensions for texture characterization and used a large sample population of all Gleason grades, but they do not provide results on the individual accuracies for distinguishing between grades 3 and 4, whereas the work in [9] only attempted to discriminate between low-grade and high-grade tumors. In previous work with texture features [10], we found that texton-based approaches, although quite successful for the Gleason grading problem, focus on the statistics of local discriminatory patterns such as nuclei abnormality. They fail to capture the larger scale cell network structures discovered by our network cycles, despite the use of multiscale filter banks. Our current approach is most similar to the work in [7, 6] that uses both texture and architectural (network) features. However, both these papers used fewer grade 3 and grade 4 samples for cross-validation. Moreover, the work in [7, 6] focused on the use of statistics based only on the triangles within the Delaunay triangulation or the polygons within the Voronoi diagram and not based upon more complex higher-order cycles. Therefore, their network features might be unable to completely capture the global nature of the discriminatory patterns between the different Gleason grades. We have provided a comparison to the methods in [10, 8] in Sec. 3.

We note that accurate cell detection is not the focus of this paper and we demonstrate excellent Gleason grading performance of our network cycle features despite imperfect cell detection. An alternative unsupervised Hessian-based filtering technique has been used for cell detection in [11]. Our method is similar in computational complexity to this technique. We note that on account of being a supervised approach, our method's accuracy could be improved with a sufficient number of expert annotations. Currently we have only trained it on relatively few non-expert nuclei/non-nuclei annotations.

2.3. Network Features

In addition to our novel network cycle features described below, we computed several well-known network features from nearest-neighbor graphs: (1) Number of vertices (equivalent to cell density since each image covers the same field of view), (2) Number of graph components, (3) Clustering coefficients, (4) Fiedler values computed from the vertex Laplacian and the edge Laplacian and (5) Average shortest path length. Some of these well-known network features are described in greater detail in [12, 5]. We shall show in Sec. 3, that for the Gleason grading task, our novel network cycle features prove to be more efficient than these other well-known features.

Our novel network cycle features first require the extraction of cycles from a network. For non-planar graphs, e.g., nearest neighbor graphs for our cell networks, one can compute a cycle basis that relies upon the initial computation of a spanning tree. The cycle basis can also be used to compute quantities such as the edge Laplacian and its Fiedler value [12]. For planar graphs such as Delaunay triangulations, network cycles can be rapidly extracted using the face-tracing algorithm in [13]. To obtain a characterization of non-triangular cycles, we remove all long Delaunay edges with

Euclidean length above a cut-off value σ. Given the different cycles, we compute their unweighted and weighted lengths (with weights based upon the Euclidean distance) and then compute various statistics: (1) number of cycles with length greater than 3, i.e., non-triangular (NT) cycles, (2) average NT cycle length, and (3) maximum NT cycle length.

3. RESULTS AND DISCUSSION

Our prostate cancer dataset consisted of 25 H&E images of Gleason grade 3 and 50 images of Gleason grade 4. Each image was acquired at 10X resolution with 0.625 micron pixel size and was of size 1392×1040 pixels. (We note that a few of our grade 4 images contained small regions belonging to other tumor grades, but we still assigned the grade 4 label to these images.)

For training our cell detector, the images were downsampled by a factor of 2, and 240 nuclei and 265 non-nuclei annotations were manually obtained. A 11×11 patch of R, G, B values was obtained from each annotation for training a linear SVM and in order to compute the sliding window outputs during the testing phase, the SVM's $11 \times 11 \times 3$ hyperplane normal coefficients were used as correlation kernels for the R, G, B channels. We use a tolerance radius of 5 during non-max suppression.

We experimented with network features computed from three different cell networks: (1) Each cell being connected to K=7 nearest neighbors (KNN), that led to a single connnected component in all 75 images, (2) ϵNN with connections between cells when the Euclidean distance between them was below $\epsilon = 15$, which led to multiple connected components in several images, and (3) Delaunay networks (σDN) obtained by removing edges longer than $\sigma = 20$ from Delaunay triangulations. No attempts were made to systematically optimize K, ϵ or σ.

For training our Gleason grading system, we selected 30 images (15 of each grade) and used the remaining 45 images for testing. We trained a decision stump classifier on each network feature described in Sec. 2.3 and examined its training and test accuracy (as percentages). The results are displayed in Table 1. It is clear from Table 1 that our network cycle features (indices 1, 4) followed by the cell density are the most discriminatory features.

Since we wish to investigate the discriminatory power of our network cycle features alone, we also used the promising orthogonal combination of unweighted and weighted average cycle lengths (feature indices 2,4) with an SVM classifier. We used the SVM with the standard radial basis function kernel ($\gamma = 1/numberFeatures, C = 1$) operating on the selected network features after Z-score normalization. We display the SVM classification boundary contours in this 2-D feature space in Fig. 3, indicating perfect separability.

Our classification results on the test dataset of 45 images are displayed in the first half of Table 2. We have tabulated the classifier accuracy in correctly identifying each tumor grade separately, the overall classifier accuracy as well as the area under the ROC curve (AUC). Note that the AUC $\in [0, 1]$ is immune to the classifier trade-off between grade 3 accuracy and grade 4 accuracy. In addition to our network cycle features, results with the texton-based and fractal-based systems described in [10, 8], respectively, are also included, neither of which can deliver the perfect separability obtained by our network cycle features. In order to avoid any bias from our training/testing partition, we also run all algorithms on 10 different train/test partitions (still with 15 samples per class in the training set) and have displayed the mean test results in the second half of Table 2. Network cycle features continued to outperform fractal and texton features in terms of mean AUC values, although for some partitions, a poor automatic choice of the SVM threshold from the training samples led to sub-optimal Grade4 classification accuracy.

For the approach in [8], we computed the 8 recommended

Table 1. Training / testing accuracy for individual network features (Please see text for explanation)

Feat. No.	Feature Description	Net Train Acc.	Net Test Acc.
1.	σDN: No. NT cycles	93.33	97.78
2.	σDN: Mean NT cycle len.	80.00	68.89
3.	σDN: Max. NT cycle len.	86.67	64.44
4.	σDN: Mean wt. NT cycle len.	100.00	95.56
5.	σDN: Max. wt. NT cycle len.	60.00	64.44
6.	No. vertices (cell density)	93.33	93.33
7.	KNN: Clustering coefficient	83.33	51.11
8.	KNN: Vertex Fiedler value	73.33	62.22
9.	KNN: Edge Fiedler value	63.33	42.22
10.	ϵNN: No. network components	83.33	80.00
11.	ϵNN: Mean Comp. Path Len.	90.00	68.89

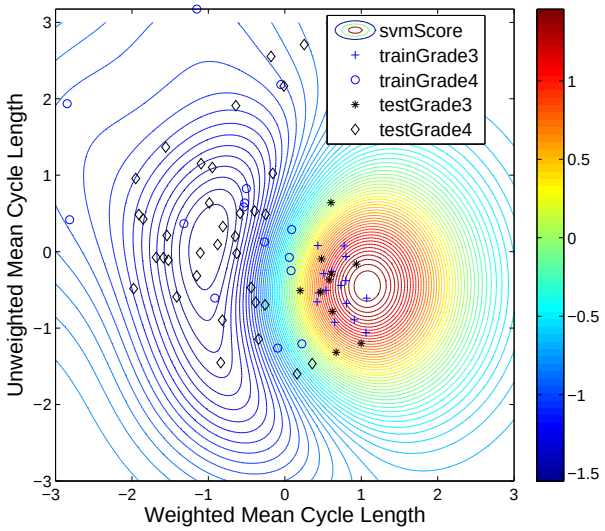


Fig. 3. SVM boundary contours on network cycle features

Table 2. Test classification results using the single training/testing partition in Fig. 3 and average of 10 such partitions

	Gr. 3 acc.	Gr. 4 acc.	Net acc.	AUC
1 train/test partition:				
CycleMeans, RBF-SVM	100.00	100.00	100.00	**1.000**
KmeansTextons, RBF-SVM[10]	91.00 $\pm$3.16	95.42 $\pm$1.47	94.44 $\pm$1.17	0.987 $\pm$0.006
Fractals, RBF-SVM [8]	100.00	91.43	93.33	1.000
10 train/test partitions:				
CycleMeans, RBF-SVM	100.00 $\pm$0.00	89.14 $\pm$4.82	91.56 $\pm$3.75	**0.995** $\pm$0.006
KmeansTextons, RBF-SVM [10]	94.00 $\pm$7.00	93.14 $\pm$7.40	93.33 $\pm$5.13	0.987 $\pm$0.002
Fractals, RBF-SVM [8]	85.00 $\pm$9.72	79.43 $\pm$4.63	80.67 $\pm$3.48	0.934 $\pm$0.035

fractal dimensions computed from gray-scale converted images rather than the recommended R-channel since this led to greater accuracy on our dataset. Note that we have only compared with the baseline K-means approach in [10] (with 8 clusters per class) since the other texton generation techniques only reduce training time without significantly affecting accuracy. However, our texton-based AUCs are an improvement over the results in [10] since we did not use color information in that paper. As in [10], to account for the randomized pixel samples picked during texton training, we obtained error bars on the test performance by using multiple training/testing runs (10) with the single train/test partition. Combination of our pre-selected network cycle network features with the remaining network features or of the pre-selected network-based features with the texton or fractal features did not lead to improvement in accuracy (and hence these results have been omitted from Table 2), but this conclusion might change with more samples.

4. CONCLUSION

Our two novel network cycle features have yielded perfect separability between Gleason grades 3 and 4 on an unseen dataset consisting of 45 images, a sample size comparable with the largest used in the existing literature for this problem. In future work, we plan to validate our conclusions with more samples and to train our system on additional tumor grades, stroma, benign epithelium and to then use our automatic Gleason grading system on whole-mount histopathology slides. Given more samples, it might prove advantageous to combine our texton features [10] with the network cycle features since the texton features might contain appearance information complementary to the cycle structure. In addition, we also plan to train our system on near-infrared images, in addition to H&E images, and to use our classifiers for distinguishing between PIN (Prostatic Intraepithelial Neoplasia) and BPH (Benign Prostatic Hyperplasia). We also plan to explore the utility of our network cycle features in other histological cancer grading applications.

5. REFERENCES

[1] Tobias Franiel, Lutz Ludemann, Birgit Rudolph, Hagen Rehbein, Carsten Stephan, Matthias Taupitz, and Dirk Beyersdorff, "Prostate MR imaging: Tissue characterization with pharmacokinetic volume and blood flow parameters and correlation with histologic parameters," *Radiology*, vol. 252, no. 1, pp. 101–108, 2009.

[2] P. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Modern Pathology*, pp. 292–306, 2004.

[3] J. Meyer et al., "Breast carcinoma malignancy grading by Bloom Richardson system vs. proliferation index: reproducibility of grade and advantages of proliferation index," *Modern Pathology*, pp. 1067–1078, 2005.

[4] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001.

[5] C. Gunduz, B. Yener, and S. Gultekin, "The cell graphs of cancer," *Bioinformatics*, vol. 20, no. Supp. 1, pp. 145–151, 2004.

[6] Scott Doyle, Mark Hwang, Kinsuk Shah, Anant Madabhushi, Michael D. Feldman, and John E. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *ISBI*, 2007, pp. 1284–1287.

[7] Shivang Naik, Scott Doyle, Shannon Agner, Anant Madabhushi, Michael D. Feldman, and John Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *ISBI*, 2008, pp. 284–287.

[8] Po-Whei Huang and Cheng-Hsiung Lee, "Automatic classification for pathological prostate images based on fractal analysis," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 7, pp. 1037–1050, 2009.

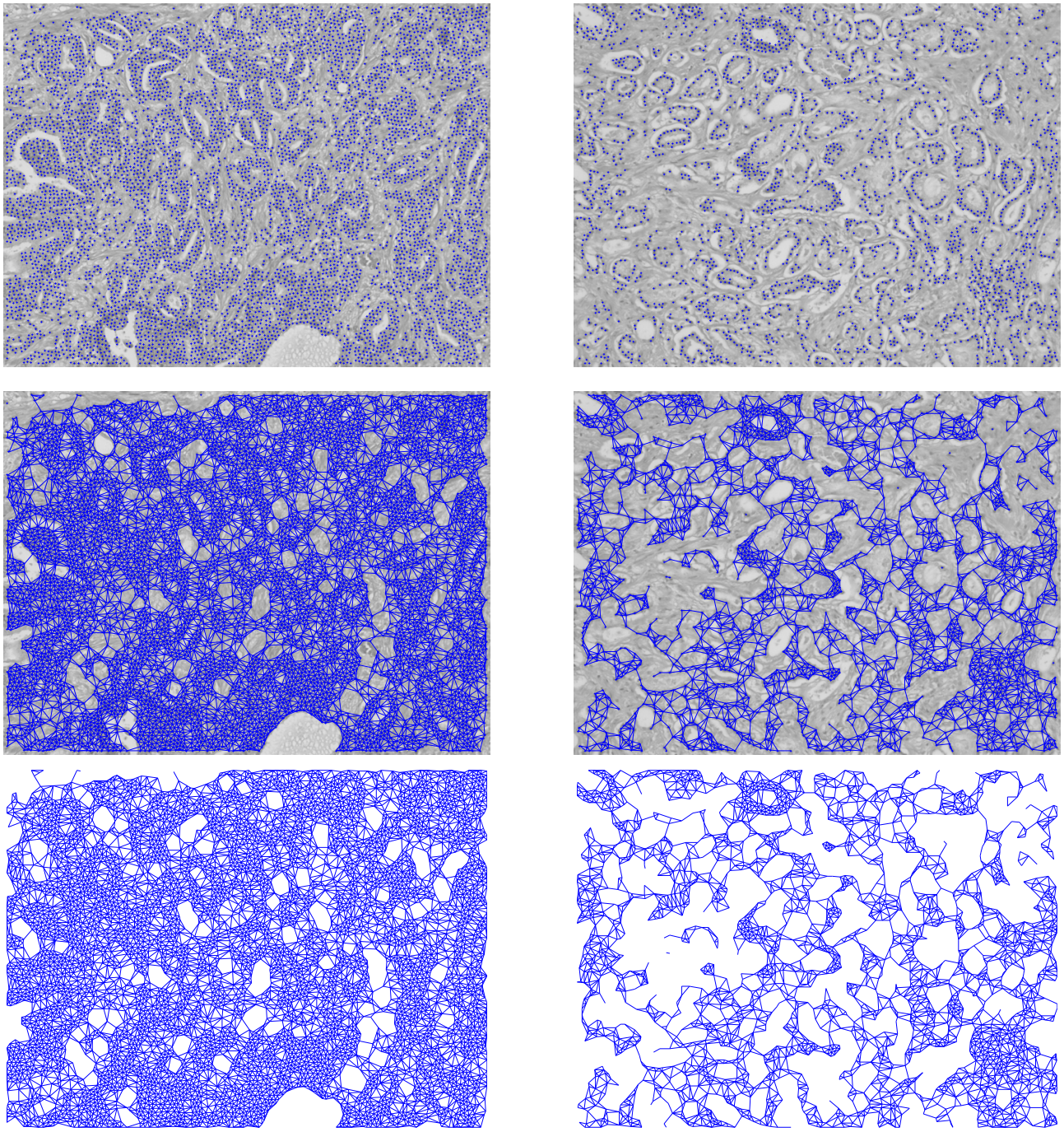
[9] A. Tabesh, M. Teverovskiy, Ho-Yuen Pang, V.P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *Medical Imaging, IEEE Transactions on*, vol. 26, no. 10, pp. 1366 – 1378, 2007.

[10] P. Khurd et al., "Computer-aided Gleason grading of prostate cancer histopathological images using texton forests," in *ISBI*, 2010, pp. 636–639.

[11] C. Bilgin, P. Bullough, G. Plopper, and B. Yener, "ECM-aware cell-graph mining for bone tissue modeling and classification," *Data Mining and Knowledge Discovery*, vol. 20, no. 3, pp. 416–438, 2010.

[12] L. Grady and J. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*, Springer, 2010.

[13] J. Gross and J. Yellen, *Graph Theory and its Applications*, Chapman-Hall, 2006.



(a) Gleason 3 (b) Gleason 4

Fig. 2. In the top two rows, cell nuclei (marked as $+$) and Delaunay triangulations (with edge lengths > 20 removed) are overlaid on the H&E images from Fig. 1, respectively. In the last row, apart from the cell density differences, note the differences in the cycle structures in the Delaunay networks.