1 Graph-Based Analysis of Cancer Cell Invasion

In this chapter a graph-based approach of studying cell invasion is introduced. The starting point were experiments performed by the group of Peter Friedl (Radboud University, Nijmegen), who measured the invasion of irradiated melanoma-MV3 cells. More specifically, they conducted measurements with a variety of gene knock-downs (KDs), where the expression of certain selected genes was suppressed. The objective was and is to find genes, which are responsible for the observed invasion-associated resistance towards radiotherapy and the formation of collective cell niches (Fig.1.1).

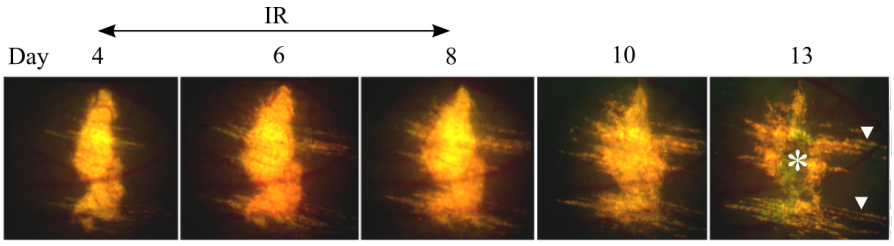

Figure 1.1: **Invasion-associated radiotherapy-resistance of a tumor** Shown are epifluorescence images at various time steps of an irradiated (IR) tumor (HT-1080 fibrosarcoma cells), which invaded into the surrounding three-dimensional collagen matrix. The main observation was that cells seemed to regress in the tumor core (asteriks) but remained persistent in the invasion zone (arrowheads). (Data and images taken from Fig.S2 of [1].)

1.1 Constructing the Graph

The *in-vitro* cell-invasion assay was embedded at day 0 into a collagen gel. After 2 days the sample got irradiated with a dose of 4Gy. At day 7 the early irradiation response was measured (Fig.1.2).

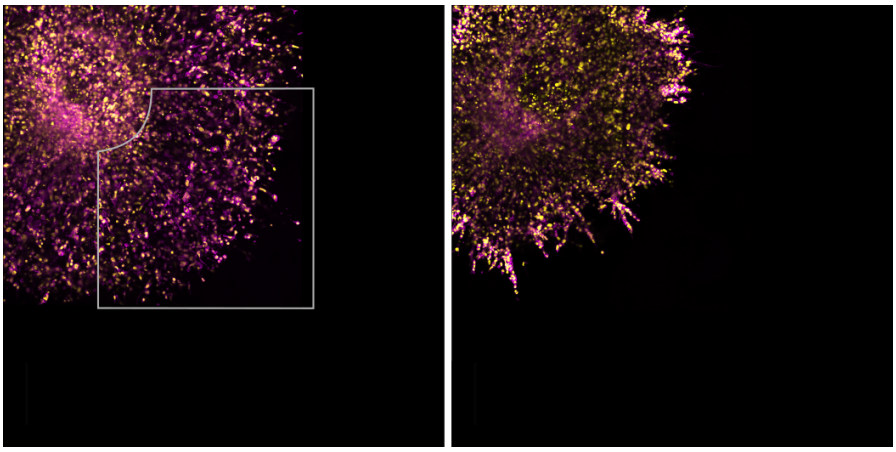

Figure 1.2: **Reflection images of invading MV3 cells**. Shown are images created through a fluorescences microscope of a two-dimensional slice of an invasion-assay of MV3 cells, where no gene was knocked-down (left) and where the expression of a specific gene (FOXG1) was reduced (right), 7 days after embedding the spheroid of cells into a collagen I gel. The cell bodies are stained violet and the cell nuclei are marked yellow. The gray area outlines part of the invasion zone. (Image credit: Peter Friedl, Radboud University, Nijmegen)

From the two-dimensional images it was possible to isolate the position of each stained nuclei, which were recorded, resulting in a set of points V corresponding to the two-dimensional coordinates of every nuclei. Importantly, it was not possible to get this information about cells within the tumor core, as cells were to clustered, making it impossible to distinguish between single cells. Therefore the number of points $|V|$ corresponds to the number of cells found within the invasion zone I with

$$I = \{\mathbf{r} \in \mathbb{R}^2 : r_0 \leq ||r|| \leq R\}, \tag{1.1}$$

where $||\cdot||$ is the Euclidean norm, r_0 is the radius of the tumor core and R is the maximal distance of a single vertex to the center of the tumor, whose position is set to the origin. The simplest way of constructing a graph from this set of points (vertices) is to construct a set of edges $E \subseteq V \times V$, where two vertices $p, q \in V$ get connected with an edge if and only if their Euclidean distance form each other is less than or equal to a distance $r \in (0, r_{max}]$, where r_{max} is defined as the upper boundary with

$$d(p, q) \leq r_{max} \;\; \forall p, q \in V. \tag{1.2}$$

For our specific case r is chosen to represent the diameter of a MV3 cells which has been reported in [2] to be approximately $15\mu m$ (Fig.1.3B). Therefore if two vertices are connected, they represent two biological cells which are in physical contact with each other. An additional constrained is added to exclude loops which would connect a vertex with itself. The construction of this cell graph (CG) follows the definition of a random geometric graph (RGG). However, in comparison to the RGG the vertices are not chosen randomly, but represent the non-uniform distribution of cells within the invasion zone. The undirected CG is fully characterized by $G = (V, E)$.

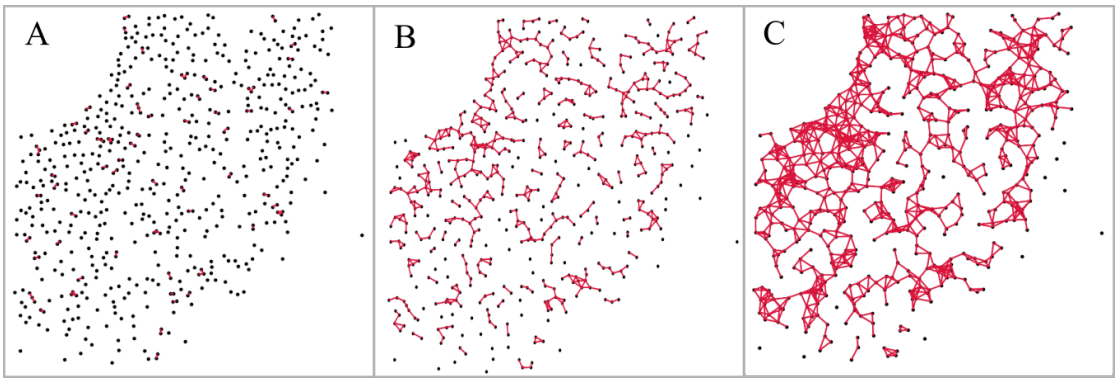

Figure 1.3: **Cell Graph of invading MV3 cells** Shown is the CG corresponding to the gray area of experimental data shown in Fig.1.2. The vertices (positions of the cell nuclei) are shown as black dots. The graph is constructed in a way that two vertices get connected with an edge if there distance is less than or equal to a parameter r. If r is small (A, $r = 5\mu m$), the largest connected component will be close to $\tilde{L} \approx 1/|V|$. Upon increasing the value r (B, $r = 15\mu m$) the largest connected component increases and multiple large connected components exist. For high values of r (C, $r = 30\mu m$) only one spanning cluster exists with $\tilde{L} \approx 1$

.

1.2 Metrics of the Graph

A similar approach of using a simple geometric graph, where the vertices are given by cell positions and edges are added depending on their distance, was reported in [3]. There the graph-based approach was used to differentiate and distinguish between cancerous and healthy or inflamed tissue by analyzing two metrics of the graph, which describe local organization. Namely the degree of each vertex and the clustering coefficient were evaluated. Based on this previous analysis showing that these two metrics can reproduce key features of cancerous tissue, they are also used in the analysis of the CG to study different invasive phenotypes under irradiation. Additionally, a metric of connectedness is introduced, which describes whether or not vertices are connected on a non-local scale via a path.

1.2.1 Vertex Degree

The degree of a vertex $q \in V$ is given by the number of edges, which are connected to it. The sum over all vertex degrees is given by the degree sum formula [4]

$$\sum_{p \in V} deg(p) = 2|E|. \tag{1.3}$$

A global vertex degree $\tilde{D}$ can be defined as

$$\tilde{D} = \frac{\sum_{p \in V} deg(p)}{|V|}, \tag{1.4}$$

$$= \frac{2|E|}{|V|}, \tag{1.5}$$

which is a measure of how many neighbors are on average connected by an edge to each vertex.

1.2.2 Cluster Coefficient

Another metric, which is used to study collective behavior within different networks is the global clustering coefficient [5]

$$\tilde{C}_{cluster} = \frac{1}{|V|} \sum_{p \in V} c_{cluster,p}, \tag{1.6}$$

with the local clustering coefficient of a vertex $p \in V$ given by

$$c_{cluster,p} = \frac{2|E_p|}{k_p(k_p - 1)}, \tag{1.7}$$

where k_p describes the number of neighbors and $|E_p|$ counts the number edges between neighbors. This can be understood as a measure of how connected the neighbors of each vertex are by comparing the number of total links between them to the maximum number of links $\frac{k(k-1)}{2}$ (in an undirected graph). The global clustering coefficient is a measure of connectivity within the direct neighborhood of each vertex.

1.2.3 Giant Component

In graph theory two vertices $p, q \in V$ of a graph G are call connected if and only if there exists a path in G from p to q. A path is defined as sequence of edges, which joints a sequence of vertices [6]. Connectivity can be used to describe an equivalence relation $\sim$ over the set of vertices V, as it fulfills all conditions of reflexivity, symmetry and transitivity for all vertices $p, q, r \in V$ with

$$p \sim p, \tag{1.8}$$

$$p \sim q \rightarrow q \sim p, \tag{1.9}$$

$$p \sim q \wedge q \sim r \rightarrow p \sim r. \tag{1.10}$$

The equivalence class of a vertex p can be expressed as the set

$$[p] \coloneqq \{q \in V : p \sim q\}, \tag{1.11}$$

which can be used to describe a local connectedness $c_{connect}$ with

$$c_{connect,p} \coloneqq |[p]|, \tag{1.12}$$

which describes how many vertices are connected via a path to the vertex p. The sum of all sets

$$P = \{p \in V : [p]\}, \tag{1.13}$$

is now used to partition the set V into connected components (subgraphs), which are disjoint as each vertex is member of exactly one connected component. The largest connected component (giant component) $|L|$ can be defined as the cardinality of the largest subgraph with

$$L \in P : |L| \geq |X| \quad \forall X \in P. \tag{1.14}$$

Similarly to the node degree and clustering coefficient it is useful to normalize the giant component

$$\tilde{L} \coloneqq \frac{|L|}{|V|}. \tag{1.15}$$

The introduced global metrics $M = \{\tilde{D}, \tilde{C}, \tilde{L}\}$ and local metrics $m = \{deg, c_{cluster}, c_{connect}\}$ will in the following be used to analyze different CGs. As it is complicated to envision how these metrics are calculated on complex graphs with high amount of vertices, two simple graphs and their metrics are shown in Fig.1.4.

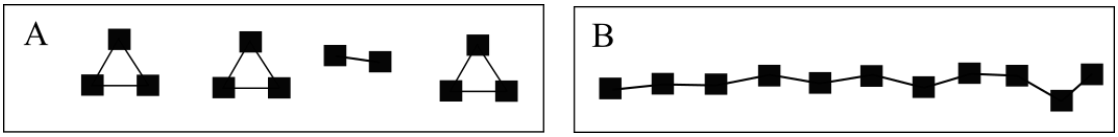

Figure 1.4: **Simplified example graphs.** Shown are two simple undirected example graphs. In both cases the number of edges and vertices is equal with $|V_A| = |V_B| = 11$ and $|E_A| = |E_B| = 10$. Therefore the global vertex degree is also the same with $\tilde{D}_A = \tilde{D}_B = \frac{20}{11}$. The clustering coefficient for graph A is much larger, with $\tilde{C}_A = \frac{9}{11}$ than for example B with $\tilde{C}_B = \frac{0}{11} = 0$, because in B there are no connections between neighbors. The giant components are given as $\tilde{L}_A = 3/11$ and $\tilde{L}_A = 1$

1.3 Cell Graphs of Irradiated Tumor Invasion

The aim of the research is to study the impact of different gene-KDs in irradiated tumors. For this reason a broad range of different gene-KDs, which included among others AOX1, FGF1, FOXG1, and ID1, were experimentally measured with irradiation (4Gy) and without irradiation (0Gy). For each measurement the image (Fig.1.2) of invading cells was recorded. These images are the superposition of three two-dimensional slices, which were recorded separately with a height difference of 10μm. For the analysis each two-dimensional slice was used to construct a distinct graph. The various KDs were imaged in 3 to 20 independent measurements for both the 0Gy and 4Gy condition. This resulted in 9 to 60 different graphs depending on the KD, whose global metrics M were subsequently calculated. The untransduced spheroid was measured in 130 independent measurements.

1.3.1 Global Analysis

The global metrics M of the graphs for all analyzed knock-downs and the unaltered (untransduced) spheroids are shown in Fig.1.6 with and without irradiation. It is important to note, that because the sample size for some of the KDs was very low, all data points are shown and not just their means, as some of them would not be statistically significant. This is reasonable, as the aim of this preliminary analysis was to identify different KDs which differ from the "norm" of the untransduced measurements and not just yet calculate statistically relevant global metrics for each KD, which would require an infeasible amount of experiments. We were primarily interested in the KDs capability to change the behavior of cells when irradiated. Therefore we wanted to identify gene-KDs, which exhibit different metrics for irradiation than would be expected by looking at the same metric for the untransduced measurements. Three main observation were made by looking at the metrics of the different graphs and comparing them qualitatively.
Firstly, the giant component $\tilde{L}$ of the untransduced cells decreased significantly, as almost all graphs of the 4Gy condition exhibited a giant component spanning less than fifty percent of all vertices. In contrast some graphs of specific KDs (i.e. AOX1 or FGF1), still show giant components, which include almost all vertices. A separation can be introduced *ad hoc* by defining graphs of the 4Gy condition as high connectivity graphs if $\tilde{L} \geq \tilde{L}_{untrans,max}$ and low connectivity graphs if $\tilde{L} < \tilde{L}_{untrans,max}$. Even though this choice of separation is arbitrary, it still achieves the objective of filtering out specific KDs which deviate the strongest from the expected behavior of unaltered irradiated cells.
Secondly, this change in giant component is not caused by a change in the number of vertices $|V|$. Both parameters seem uncorrelated, meaning there are graphs with very low number vertices showing the same giant component as graphs with high number of vertices.
Lastly, the clustering coefficient $\tilde{C}$ as a function of vertex degree $\tilde{D}$ does not change for the irradiated cells compared to their untreated counter parts. This can be interpreted as the fact that cells do not change their local spatial organization in response to radiation. This also suggest, that determining either the vertex degree $\tilde{D}$ or the clustering coefficient $\tilde{C}$ is sufficient to describe local organization.
To better understand the values of these metrics and their interdependences, a simple Random Geometric Graph (RGG) is defined, where cells are assumed to be uniformly distributed through

out the invasion zone. This is achieved by randomly selecting n coordinates within the invasion zone I as the set of vertices of the RGG. The construction of edges follows the definition of the CG. A visual comparison between a CG and a RGG is shown in Fig.1.5. To quantitatively compare the metrics between the two graphs the number of vertices of the RGG, which directly determines its metrics, was varied systematically in the range of $|V_{RGG}| \in [0, |V_{CG,max}|]$. The resulting curves are included in Fig.1.6. By comparison it can be seen that the CGs are less clustered (lower clustering coefficient) but more connected (larger giant component) than the RGGs.

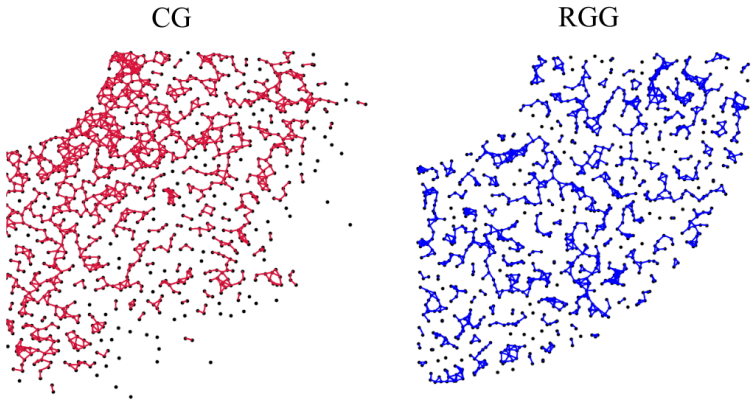

Figure 1.5: **Comparison between the Cell Graph (CG) and a Random Geometric Graph (RGG).** The nodes of the CG correspond to the position of cell nuclei, while the RG was obtained by just uniformly distributing a number of vertices into the invasion zone. In both cases the number of vertices is equal, with $|V_{CG}| = |V_{RG}| = 1040$. The global metrics for the CG are $\tilde{D} = 2.7$, $\tilde{C} = 0.3$ and $\tilde{L} = 0.5$. The global metrics for the RG are $\tilde{D} = 2.9$, $\tilde{C} = 0.5$ and $\tilde{L} = 0.1$.

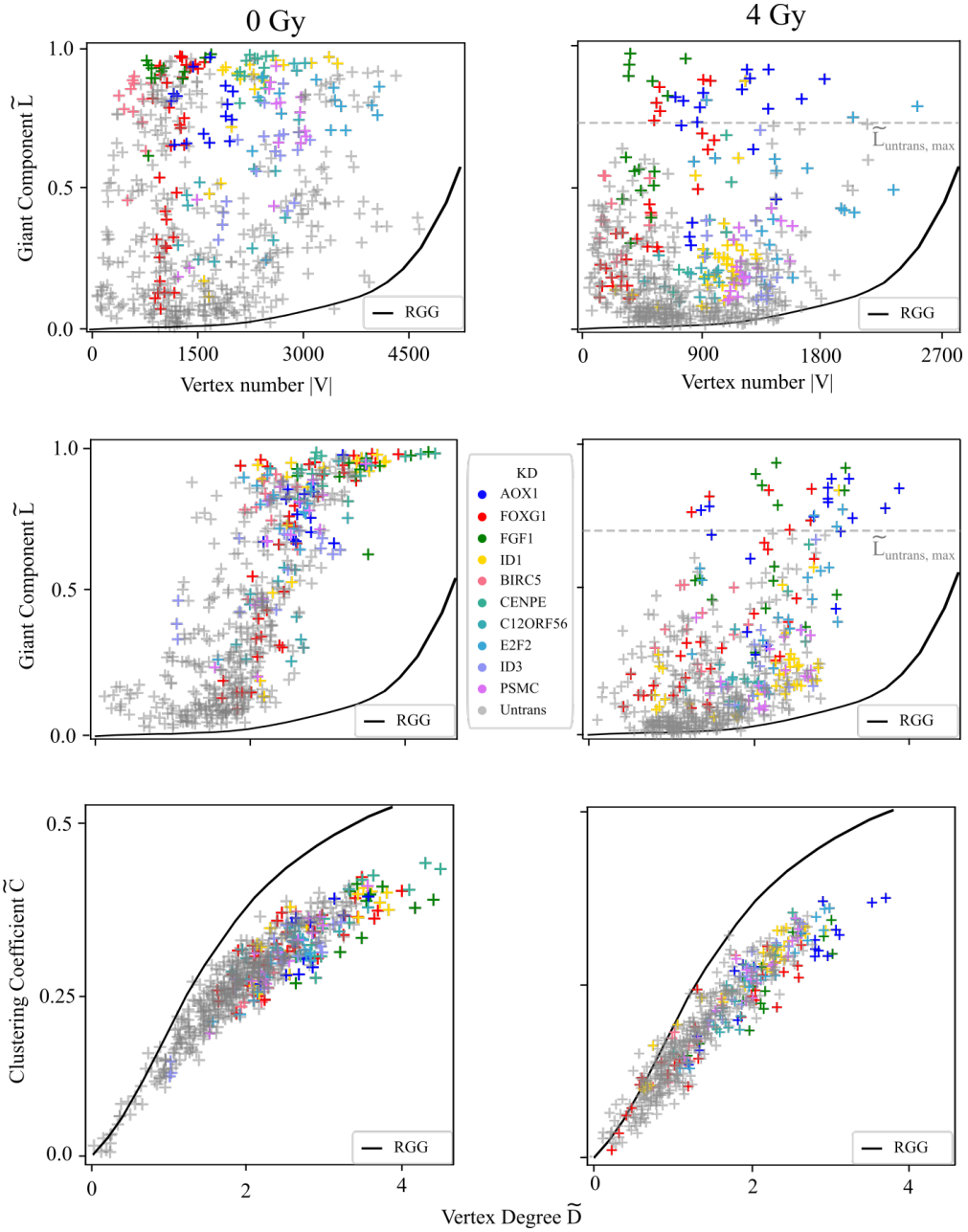

Figure 1.6: **A global metric of the CG identifies different KDs which show high connectivity between cells in the irradiated tumor.** Shown is interdependence of the global graph metrics $\tilde{D}, \tilde{C}$ and $\tilde{L}$. Additionally, the giant component $\tilde{L}$ is shown against the number of vertices of each graph $|V|$. The different graphs were constructed and analyzed for 0Gy (left) and 4Gy (right). Each point represents a graph of a single measurement, where either one specific gene (indicated by different colors) or no gene (gray) was knocked-down. The highest value for the giant component for the untransduced measurements $\tilde{L}_{max}$ is shown as a gray dotted line. Additionally, the metrics of random geometric graphs are shown (black line), which were calculated for varying number of vertices in $|V_{RGG}| \in [0, |V_{CG,max}|]$

1.3.2 Connectivity Analysis of Selected Gene-KDs

The analysis of the global metrics revealed that there are certain KDs, which showed high connectivity (large giant connected component) of the irradiated cells. As this is exactly the behavior of cells collectively evading cell death by irradiation, we wanted to study, it is interesting to analyze this subset of graphs in more detail. One way of constructing a deeper analysis suggests itself as determining how the spatial organization (local metrics) of vertices changes within the invasion zone. Because of rotational symmetry the local metrics of single vertices depend only on the distance to the core. For this reason it is useful to define a parameter m_r, which represents an average metric m for all vertices $p \in V$ with $||p|| = r$. This definition is reasonable in the limit of high number of vertices. If the number of vertices is small it is practical to expand this definition to allow vertices within a small range around r with $r - \frac{\epsilon}{2} \leq ||p|| \leq r + \frac{\epsilon}{2}$, with $\epsilon \to 0$ for $|V| \to \infty$. Geometrically, this definition can be used to partition the set of vertices V into a set of n disjoint annuli, where the area of the annulus with *mean* distance r is given by

$$A_r \coloneqq \{p \in V : r - \frac{\epsilon}{2} < ||p|| \leq r + \frac{\epsilon}{2}\}, \tag{1.16}$$

with $r \in [\frac{\epsilon}{2}, \frac{3\epsilon}{2}, ..., R_{max} - \frac{\epsilon}{2}]$ and $n\epsilon = R_{max}$. The number of annuli n was determined using the Sturges' formula [7], which approximates the width of each class interval with

$$n = \log_2 |V| + 1, \tag{1.17}$$

where $|V|$ represent the number of cells, which were measured within the invasion zone. The average local metric of all vertices with a specific distance to the core can be defined as

$$m_r \coloneqq \frac{\sum_{p \in A_r} m(p)}{|A_r|}. \tag{1.18}$$

As we are interested primarily in the changes which arise in the 4Gy compared to the 0Gy condition, it is useful to define a relative change Δm_r, which compares the values for the 4Gy condition to the maximum of the 0Gy condition with

$$\Delta m_r = \frac{m_{r,4Gy} - m_{max,0Gy}}{m_{max,0Gy}}, \tag{1.19}$$

where m_{max} is the maximum value for m_r for the graph constructed for the 0Gy condition. In simple terms this relative parameter allows to compare how much the local metrics of different KD are effected by irradiation. (Fig.1.7).

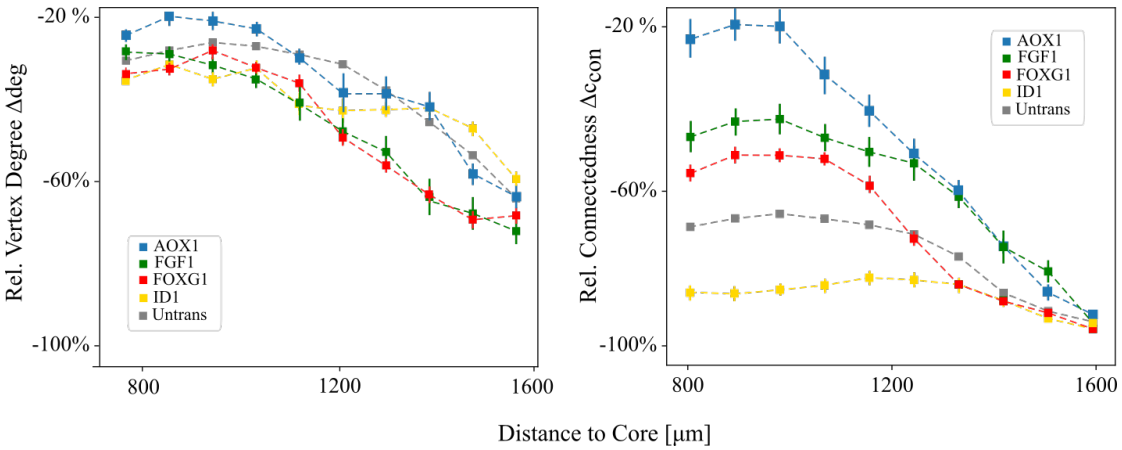

Figure 1.7: **High connectivity is not accompanied by a high vertex degree** Shown is the change in vertex degree Δdeg (A) and connectedness Δc_{con} (B) compared to their respective maximum value for the 0Gy condition against the distance to the core. Both metrics indicate that the graphs had less neighbors and were less connected than the 0Gy graphs ($\Delta deg, \Delta c_{con} < 0$). In general both parameter decrease as a function of distance, indicating that cells are less connected on a local and non-local scale. While the change in vertex degree is relatively similar for all KDs, the relative connectedness shows a substantial amount of deviation. Especially close to the core KDs of AOX1, FGF1 and FOXG1 remained more connected than the untransduced baseline, while ID1 was less connected. This is the same behavior which is also reflected in the giant components of these KDs (Fig.1.6). Interestingly, cells with KDs of FOXG1 and FGF1 show a higher decrease in vertex degree than the untransduced cells, which means that even though they are on average connected to more vertices via a path, the number of direct neighbors decreases. The average number of vertex for the different KDs was $N_{AOX1} \approx 1300$, $N_{FOXG1} \approx 550$, $N_{FGF1} \approx 550$, $N_{ID1} \approx 1200$ and $N_{untrans} \approx 850$. From Sturges' formula the number of data points n on the x-axis representing the number of intervals was determined as $n \in \{10, 11\}$. For simplicity the number of intervals was set to 10 for all KDs, this implies some over smoothing of the data of AOX1 and ID1.

1.4 Results and Discussion

Here a method of abstracting an image of cell invasion data into a graph was presented. Through the definition of a set of suitable metrics on the graph, it was possible to identify different gene KDs, whose graphs exhibited a large connected component after irradiation, which is not expected if there is no gene-KD. Biologically this means, that there is a group of KDs, whose cells withstand the damage caused by irradiation by strongly connecting with each other. Interestingly, this increased connectivity is accompanied by a lower amount of direct neighbors (for FOXG1 and FGF1). This can be interpreted as the fact, that cells form elongated cell-chain like structures instead of clusters in response to irradiation. A hypothesis about the underlying mechanics, which is to be verified experimentally, is that cells try to maximize their surface area and thereby their integrin bonds with the surrounding ECM, while still being connected to other cells. This suggests that both cell-ECM as well as cell-cell adhesions cooperate in a signaling process to up-regulate the ability of cells to normalize the damage caused by irradiation.
This shows how a purely geometric approach of constructing a graph can help to infer informations about the underlying biological processes of invading cells.
Besides the information that this graph-based approach yields to formulate testable hypothesis, the local and global metrics of the cell graphs can also be used to validate mathematical models. More specifically, it will be part of future research to use Bayesian-model selection as well as parameter inference schemes [8] to infer which model type and variant can best explain the collective effect of invasion and therapy-resistance.

Bibliography

[1] A. Haeger, S. Alexander, M. Vullings, F. M. Kaiser, C. Veelken, U. Flucke, G. E. Koehl, M. Hirschberg, M. Flentje, R. M. Hoffman, *et al.*, "Collective cancer invasion forms an integrin-dependent radioresistant niche," *J. of Exp. Med.*, vol. 217, no. 1, 2020.

[2] M. Pødenphant, N. Ashley, K. Koprowska, K. U. Mir, M. Zalkovskij, B. Bilenberg, W. Bodmer, A. Kristensen, and R. Marie, "Separation of cancer cells from white blood cells by pinched flow fractionation," *Lab on a Chip*, vol. 15, no. 24, pp. 4598–4606, 2015.

[3] C. Gunduz, B. Yener, and S. H. Gultekin, "The cell graphs of cancer," *Bioinf.*, vol. 20, no. suppl_1, pp. i145–i151, 2004.

[4] D. S. Gunderson and K. H. Rosen, *Handbook of mathematical induction*. CRC Press LLC, 2010.

[5] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[6] J. A. Bondy, U. S. R. Murty, *et al.*, *Graph theory with applications*, vol. 290. Macmillan London, 1976.

[7] H. A. Sturges, "The choice of a class interval," *J. of the Am. Stat. Asso.*, vol. 21, no. 153, pp. 65–66, 1926.

[8] E. Klinger, D. Rickert, and J. Hasenauer, "pyabc: distributed, likelihood-free inference," *Bioinf.*, vol. 34, no. 20, pp. 3591–3593, 2018.