Ice Lead Network Analysis

Julia Kaltenborn
McGill University & Mila
Montréal, Canada
julia.kaltenborn@mail.mcgill.ca

Venkatesh Ramesh
Université de Montréal & Mila
Montréal, Canada
venkatesh.ramesh@umontreal.ca

Thomas Wright
McGill University & Mila
Montréal, Canada
thomas.wright2@mail.mcgill.ca

ABSTRACT

Ice lead analysis is an essential task for evaluating climate change processes in the Arctic. Ice leads are narrow cracks in the sea-ice, which build a complex network [23]. While detecting and modeling ice leads has been done in numerous ways based on airborne images [2, 6, 7, 13, 15, 17, 20, 28], the dynamics of ice leads over time remain hidden [12] and largely unexplored. These dynamics could be analyzed by interpreting the ice leads as more than just airborne images, but as what they really are: a dynamic network. To this end, we compiled a new dataset that contains dynamic ice lead graphs. The data is based on the dataset by Hoffman et al. [6] which contains daily ice lead observations from Moderate Resolution Imaging Spectroradiometer (MODIS) between 2002 and 2020. In this project, we conduct a spatio-temporal analysis of the ice leads with the newly created graph data, exhibiting seasonal and annual trends in the ice lead dynamics. We further perform a network analysis of the ice lead graphs, which exhibit unique characteristics that diverge from those present in common real-world networks. This new perspective of interpreting ice lead as networks reveals that challenges such as ice lead forecasting and tracking might be feasible with the right network science tools. However, current network science methods, ranging from preferential attachment to EvolveGCN, are not suitable for these tasks due to the distinctive structure of the ice lead networks. From a network science perspective, the ice lead networks represent a promising case study for the development of more generalizable algorithms. From a cryospheric science perspective, the interpretation of ice leads as networks presents a unique and new opportunity to track and forecast ice leads more efficiently. This work reveals the potential of interpreting ice leads as networks and is a call for extending current network analysis methods for a new class of real-world dynamic networks.

KEYWORDS

ice leads, networks, dynamic graphs, graph neural networks

ACM Reference Format:
Julia Kaltenborn, Venkatesh Ramesh, and Thomas Wright. 2021. Ice Lead Network Analysis. In *Placeholder City '21: ACM Symposium on Fancy Research Area, June 00–99, 2021, Placeholder City, State.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/placeholder

Placeholder City '21, June 00–99, 2021, Placeholder City, State

ACM ISBN 978-1-4503-XXXX-X/placeholder…$0.00
https://doi.org/10.1145/placeholder

1 INTRODUCTION AND MOTIVATION

The Arctic climate and sea-ice condition have strong feedback effects on global climate [26], making the analysis of Arctic ice dynamics key for climate change mitigation and adaptation. One important factor in Arctic climate and sea-ice loss are ice leads since they are massively involved in Arctic melting and climate dynamics.

Ice leads are narrow, elongated cracks in the sea-ice that branch and intersect to build complex networks [23]. While only 1 - 2 % of the central Arctic is covered by ice leads, they are responsible for large heat releases from the Arctic ocean into the atmosphere, driving more than 70 % of all upward heat fluxes [11]. The large temperature and moisture gradients between air and water cause these large heat exchanges, and in summer, the low albedo of open water (<0.1) reinforces this effect further [11].[1] These heat exchanges are a major driver of sea-ice loss, to the degree with which we can even predict sea-ice loss with ice leads [29]. Due to their involvement in Arctic energy fluxes, ice lead detection and modeling are highly relevant tasks for polar scientists, as described further in Section 2.

A major aim within the ice lead research field is to understand the dynamics of ice leads, i.e. their spatial and temporal behavior, as well as their interaction with changing climate conditions. Arctic climate models require detailed information about ice lead behavior to simulate ocean-atmosphere interactions accurately [21]. Since ice lead dynamics remain largely an open question [12], we propose a new approach to analyze ice lead dynamics: a network analysis of ice leads. The network perspective is a simplified, faster alternative to current physical ice lead models and might enable, with the right tools, lead branch forecasting and tracking.

Our results show that a spatio-temporal analysis of the ice lead networks is not only possible but revealing. That said, tackling a task such as ice lead forecasting will only become possible with the adaptation of current network science methods for less usual network structures. We unexpectedly discovered that the ice lead networks pose some uncommon network characteristics, including but not limited to planarity, disconnectivity, and high temporal variability. These characteristics make e.g. many common link prediction methods unsuitable for this type of real-world network. For this reason, we present the ice lead network dataset to the network science community to enable the study of this type of network and the development of new, more generalized methods.

The contributions of this work are:

- Provide an ice lead network dataset based on the MODIS observations from Hoffman et al. [6]
- Spatio-temporal analysis of the ice lead networks
- Network analysis of the ice lead networks

[1] A low albedo implies that more short-wave radiation from the sun is absorbed, causing heating of the water masses.

The concrete tasks this research project tackles are described in Section 3. The Hoffman et al. [6] dataset on which the network data builds upon is described in Section 4. Methods and experimental setup used to perform the ice lead and network analyses are explained in Section 5 and Section 6 and the respective results and their discussion can be found in Section 7. Final thoughts are included in Section 8.

2 RELATED WORK

Related work for this project includes works from ice lead detection, analysis, and modeling, as well as the discovery of unusual networks in the network science domain.

Prior to analyzing ice leads they first of all need to be detected on airborne images. A lot of research has been done on ice lead detection [2, 6, 7, 13, 17, 20, 28]. Almost each of the cited examples used a different measuring technique, ranging from airborne lidar systems [20] to CryoSat2 [28]. The distinct measuring methods result in very different spatial and temporal resolutions, and preprocessing algorithms, e.g. cloud masking is not always needed. Additionally, further methods to segment and finally extract the ice lead objects are necessary. Only after this heavy post-processing, it becomes possible to analyze ice lead distributions.

Since the different detection algorithms need different subsequent analyses, only a few works have actually pursued the spatio-temporal analysis of ice leads [21]. The network science approach adds to the few work on temporal and spatial distribution analysis of ice leads, providing additional analysis of the interactive behavior of ice leads. In previous work, and also more generally, the analysis of the ice leads can only be achieved for one type of dataset and cannot be re-applied to datasets with a different detection method. Our approach of network analysis could be applied to any kind of measurement data as long as it is pre-processed and provides summarized characteristics of the edges.

Other approaches to reveal the dynamics of ice lead exists, namely tracking algorithms [7] and Arctic sea-ice models [15]. There are only few ice lead tracking algorithms, but Hutter et al. [7] showed that ice lead tracking is possible when provided with additional sea ice drift information - even without extensive Arctic sea-ice models. However, tracking reveals only the spatial part of ice lead dynamics and lack predictive capabilities. In contrast, the modeling algorithms are able to provide an extensive amount of information about ice lead dynamics and are also used for forecasting [15].[2] However, these physical models are not only modeling ice leads, they have to model the complete ice dynamics taking into account many atmospheric and oceanic variables. A far simpler model encoding the dynamics of ice leads might often suffice to solve challenges such as ice lead forecasting and would be - most importantly - much faster and computationally less intense than a physical model. The network science perspective could provide the basis to develop simple and fast approaches that still have the predictive capabilities to address challenges such as ice lead forecasting.

Repeatedly, new kinds of networks [1] exhibiting uncommon features [3] and unexpected behaviors of graphs [10] have been introduced to the network science community. While the single properties of the ice lead graphs can easily be found in other networks, we have - to the best of our knowledge - not found networks with the joined properties we discovered here.

[2] There are also several other ice models, but the authors did not check them for their predictive capabilities regarding ice leads.

3 PROBLEM DEFINITION

The tasks addressed in this work can be split into (1) translating the ice leads into graphs, (2) analyzing the ice leads in terms of spatial and temporal distribution, and (3) analyzing what kind of network the ice leads build. Point (3) includes firstly, analyzing network properties such as connectivity, motifs, and more, and secondly, performing common network science tasks such as link prediction and graph classification.

The ice leads can be translated into undirected networks by interpreting the leads (i.e. the open water area) as edges in a graph, and the start/endpoints as nodes in a graph (see Figure 1). If two ice leads are crossing each other, their intersection is interpreted as a node as well. Intuitively, the interconnected black water lines visible from an ice lead picture are exactly the graph that we extract from our underlying data (Section 4). The nodes in the graph must hold the geographical locations (e.g. longitude and latitude) as node attributions and can be enriched with further location-bound information (e.g. temperature). The edges can hold information about the individual ice lead, such as width and length. Most notably, the ice leads are observed in this work over time, i.e. that the networks are *dynamic* networks with ice leads evolving over time.

The spatio-temporal analysis of the ice leads entails extracting knowledge from the provided ice lead networks. This work provides a detailed overview of spatial distributions, ice lead width and length distributions, and a temporal analysis, since only few spatio-temporal analyses have been provided so far. Additionally, the network science perspective enables analyzing interactive behavior between ice leads (e.g. by looking at diameters). The spatio-temporal analysis is key to extracting unexplored, yet important knowledge about ice leads from the dataset at hand.

The network analysis task aims to describe the type of network that ice leads are building. To this end the properties of the network must be analyzed, i.e. unique characteristics, as well as lacking common properties must be revealed. Not only the properties but also the applicability of common network science methods must be examined. Since link prediction and graph classification are interesting from an ice lead science perspective, several link and graph classification methods are tested on the ice lead networks. Graph classification on ice lead networks can be used to predict ice extent or seasonality from a given ice lead network. Link prediction translates to predicting where new ice leads might emerge in the next time step. The network analysis reveals why studying this type of network might be of interest to other network scientists.

To this end, the main questions addressed here are:

(1) **Translate ice leads into graph**:
Is it possible to interpret ice lead data as (dynamic) graphs?
(2) **Spatio-temporal Ice Lead Analysis**:
Can we derive properties and dynamics from ice lead graphs?
(3) **Network Analysis**:
How are these graphs different from other real-world graphs? Are current network science methods suitable for this type of network?

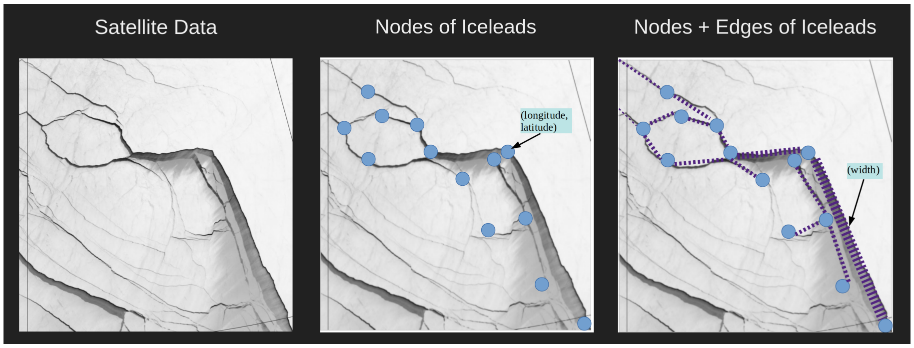

Figure 1: Interpreting ice leads as network. The first image from left shows the satellite image of an ice lead set. The second image indicates that start, end, and intersections points of ice leads are interpreted as nodes (visible as blue circles), with longitude and latitude values as node attributes. In the third image purple, dashed edges are added between the nodes where ice leads appear. Weight attributes of the edges can be e.g. the width of the ice leads. The underlying satellite image displays the Beaufort Sea on February 23, 2013 and is obtained from NASA [24].

4 DATASET DESCRIPTION

As described in the previous section and in Figure 1, the ice leads can be translated into networks by interpreting the start, end, and intersection points of ice leads as nodes, and the ice leads themselves as edges. The dynamic network can be built by joining several observations of ice leads together. In an ideal dataset, the set of nodes, edges, and their attributes can directly be extracted from the data.

We use the ice lead dataset from Hoffman et al. [6] because of its convenience for graph extraction. Numerous other data sets provide detected ice leads in image or NetCDF formats (e.g. [16]), which provide access to start, end, and intersection points of ice leads. However extracting ice leads from these datasets is very challenging: (1) The image sizes are too large to extract network objects directly from the images with python tools such as NEFI [4]. (2) The ice leads are fragmented and not joined together to the branches they actually build on a global scale, which complicates extracting the underlying network. Hoffman's dataset overcomes both of these challenges. In Hoffman's dataset, the ice lead branches are stored in txt files that are much easier to process, addressing issue (1) and their algorithm connects smaller ice lead parts to larger branches, solving problem (2). The daily observations of Hoffman's dataset contain start and endpoints of the ice leads, providing an easily accessible, initial set of nodes and edges.

Hoffman's dataset provides us also with node and edge attributes as well, however, are lacking the intersection points of ice leads. The start and endpoints of the ice leads are given in longitude and latitude values, which we use as node attributes. Provided information about the width, length [3] and open water area of the branches are used as edge attributes in our networks. Additional information, such as the region in which the ice leads were discovered is not of any concern here. Notably, the dataset does not provide intersection points of ice leads, which consequently must be calculated during our graph-translation procedure (see Section 5.1).

Hoffman's dataset contains daily ice lead observation from 2002-2020. Since the cloud coverage in the Arctic summer is comparably high to the other seasons, the dataset contains only very few examples of summer observations. The dataset can be accessed via ftp under the following address: ftp://frostbite.ssec.wisc.edu/.

5 METHODOLOGY

The methodology section is divided into three parts describing the process of translating the ice leads into graphs, analyzing the spatial and temporal distribution of ice leads, and analyzing the structure of the network and how well common network science methods can be applied to this network.

5.1 Ice Leads to Graph Translation

Depending on the dataset, different methods must be employed to translate ice lead data into graphs. For example, the direct processing of ice leads from image data requires different methods than the ones described here. However, any kind of ice lead observation can in principle be translated into the data format we use in this work, where the geographic location and important properties of ice lead branches are given or collected. As a result, the translation method described in the following extends to a majority of ice lead datasets. [4]

As a first step, naive graphs are generated from the start and endpoints of the ice lead branches collected in Hoffman's dataset [6]. The start and endpoints become a set of nodes, with longitude and latitude values as attributes. The pairings of start and endpoints that define the ice lead branches provide also the edges of the graph, with weights such as width and length of the ice lead. Notably, the graph that has been created lacks intersections between ice leads and the ability to connect very close ice leads that are e.g. partially separated by ice drift.

Intersections between ice leads can be calculated from the geographical locations of the end and start points of the ice lead

[3] actually "great circle length": length of a line a sphere, since the ice leads are on a globe

[4] This is especially the case since direct network extraction from ice lead images is challenging as described in Section 4.

branches. The pseudo-code describing how intersections are computed and added to the graph can be found in Algorithm 1. First, the given longitude and latitude values are projected onto an azimuthal equidistance coordinate system (ESRI:54032) to enable precise geometric calculations. On the projected map, all possible intersections for a set of ice leads are calculated. The identified intersections are added as new nodes into the graph. The two edges that are intersecting are deleted and replaced by four edges meeting at the intersection point. The node attributes of this newly added node are computed by back-projecting the intersection coordinates into the longitude-latitude system (EPSG:4326). The weights of the new edges are either taken over from the old weights (e.g. for width) or computed (e.g. for length). Special attention must be paid to the case where one ice lead intersects several times with other ice leads. After adding the intersection points as nodes to the graph, the node degree increases and individual ice leads have been split into sub-parts.

The graph can be further improved by summarizing close nodes together to one node. This way ice leads that are shortly disrupted by ice parts can be connected, and ice-shattered lead intersections are actually represented as intersections. This has not been incorporated due to the limited resources and restrictive timeline associated with this project.

After adding ice lead intersections to the graph, a more appropriate graph representation of the ice lead networks is obtained.

Algorithm 1 Add Ice Lead Intersections to Graph

```
Graph ← Vanilla ice-leads graph
for node = 1, 2, ..., N do
    node.X, node.Y ← Project(node.Longitude, node.Latitude)
end for
intersections ← empty list
for edge = 1, 2, ..., E do
    for next_edge = edge + 1, edge + 2, ..., N do
        intersections.Insert(if_Intersects(edge, next_edge))
    end for
end for
for intersect in intersections do
    long, lat ← BackProject(intersect.X, intersect.Y)
    graph.Add(new node, long, lat)
    new_lengths ← Calculate new lengths of edges
    graph.Add(new edges, old widths, new_lengths)
    graph.Delete(old edges)
end for
```

5.2 Ice Lead Analysis

The spatio-temporal analysis of the ice lead graphs is performed to reveal new knowledge about ice lead behavior and its change over time. We analyze the graph both from a static perspective and a dynamical perspective, i.e. we look at the changes in ice lead properties over different time horizons.

In the static analysis, the individual behavior of each graph is measured and the results are averaged over all graphs. In the dynamic analysis, these behaviors are observed in the context of time, i.e. that seasonal cycles are exposed as well as larger behavioral changes over the complete time span. The former is important in understanding how ice leads change throughout a year, exposing potentially important features for forecast models. The latter is important to analyze how the properties and the distribution of ice leads change in the context of climate change.

We use the following ice lead properties to describe the behavior of the ice lead graphs:

- Geographical location (longitude and latitude)
- Length of ice leads
- Width of ice leads
- Total water area of ice leads
- Number of ice leads (size of network)

The methods for analyzing the interactive behavior of ice leads are covered in the next Section 5.3 about network analysis. Looking at degree distributions, the behavior of graph diameters over time and connectivity values can reveal knowledge that remains hidden in the common spatio-temporal analysis.

Analyzing the properties of the ice lead network over different time horizons provides the cryospheric community with a detailed temporal analysis of ice lead behavior.

5.3 Network Analysis

The network analysis part involves both looking at general network properties and structure, as well as applying common network science methods to solve well-known tasks such as link prediction and graph classification.

5.3.1 Network Properties. In order to analyze the structure and the characteristics of the ice lead network as well as their interactive behavior the following properties are analyzed:

- Planarity
- Connectivity
- Triangles and transitivity
- Clustering coefficient
- Assortativity

Furthermore, we look at degree distributions and if the small-world phenomenon holds, as well as whether the average node distance shrinks over time - both phenomena that can be found in common (dynamic) real-world networks [8, 10].

5.3.2 Network Methods. Since the ultimate goal is to tackle a challenge such as ice lead forecasting with the help of ice lead networks, we further analyze to what extend link prediction methods can be applied to this type of network. We also touch upon graph classification, which is interesting for predicting sea ice coverage or seasons, however, the main focus lies on link prediction.

We used both a set of simple link prediction methods to operate on static graphs and we chose to use EvolveGCN [14] for a temporal setting of link prediction. In both settings, we assume that a set of edges is already known, while 5%, 10%, or 20% of the ground truth edges are dropped. The task is to predict those missing edges, indicating where ice leads are most likely to be found in the current setting.

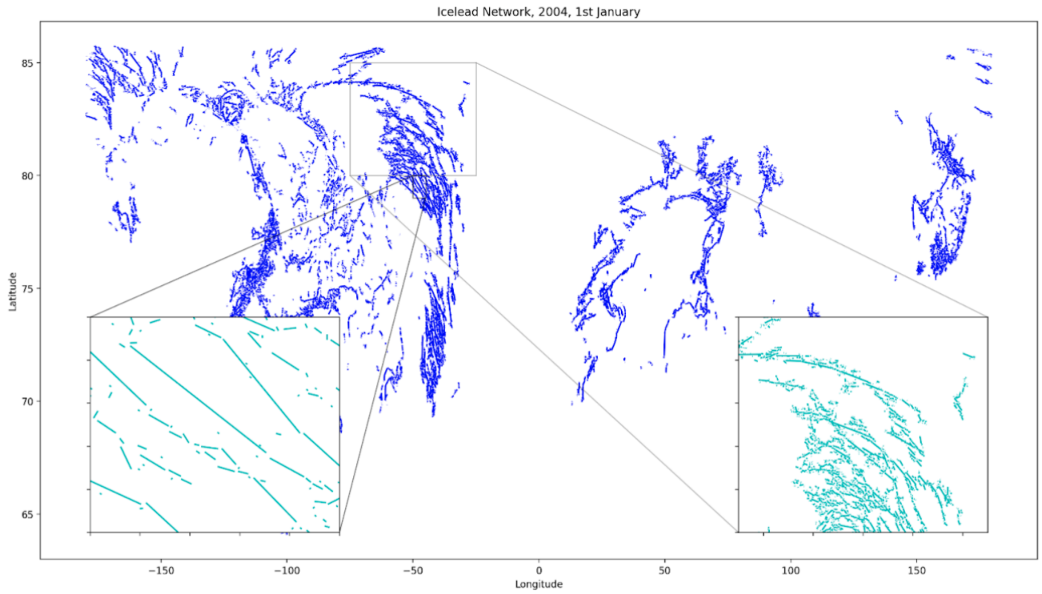

Figure 2: The ice-lead distribution for a graph from January 1st, 2014 which shows the zoomed-in network structure for some areas.

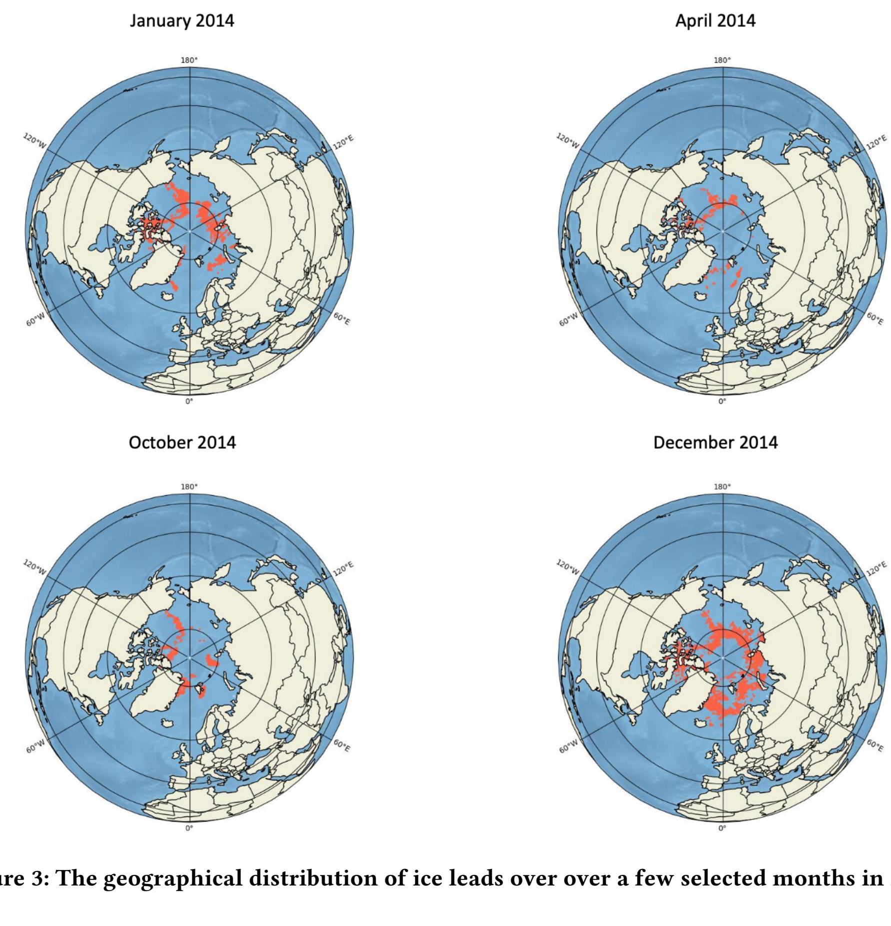

Figure 3: The geographical distribution of ice leads over over a few selected months in 2014.

For the static setting only one single, static ice lead network is considered. The following methods were applied for this setting:

- Jaccard Coefficient
- Resource Allocation Index
- Adamic Adar Index
- Preferential Attachment

All static link prediction methods available on Networkx that do no require cluster labels and were used. These methods are regarded as the "standard" methods for local similarity-based link prediction [9] and are usually employed. For the dynamic setting use only one method, namely EvolveGCN [14]. The nature of our dataset determines the set of possible methods that can be used for link prediction. Since Hoffman's dataset does not provide ice lead tracking, the resulting temporal graph can be described as a sequence of discrete graphs [22]. Thus, discrete dynamical GNNs are chosen to perform the link prediction task. We identified three different models suitable for link prediction where nodes are added and deleted in different time-steps: DySAT [19], HDGNN [30] and EvolveGCN [14]. Due to its time-spatial integrated approach and code availability, we settled for EvolveGCN to perform link prediction on the dynamic ice lead graphs.

EvolveGCN uses a recurrent neural network (RNN) to update the weights of its integrated graph convolutional network (GCN). The weights can be either regarded as hidden states or as in- and outputs of the RNN. Pareja et al.'s approach offers flexibility since the RNN and GCN can be easily replaced by alternative models (e.g. using a Long Short-Term Memory instead of the provided RNN). By using an RNN to evolve the GCN weights, EvolveGCN achieves to represent the dynamics of successive discrete graphs. [14]

We apply EvolveGCN to a link prediction setting where from a time-series of four ice lead graphs the missing edges in the last time step should be predicted. We also apply EvolveGCN for a graph classification task, where the current Arctic season should be predicted from a time-series of four ice lead graphs.

If link-prediction were possible with the described methods, this would be a major step to address the challenge of ice lead forecasting that can only be solved by complex physical models to date.

6 EXPERIMENT SETUP

The experimental setup of this project explains the environments for the graph translation and the network analysis as described in the following paragraphs. The spatio-temporal analysis requires no further specific experimental setup. Throughout the whole project, we use Python 3.9 and all experiments can be replicated with the code provided on GitHub. [5]

During the graph translation process, 3880 daily observations from 2002 - 2020 were translated into NetworkX graphs [5] and the functions provided by NetworkX were used to analyze the network properties. For the computation of ice lead intersections, the vectorized functions provided by PyGeos [25] proved helpful to reduce computing time massively. Both a parallelized and a non-parallelized version of the graph translation method exist to address different demands: The non-parallelized graph translation takes on a local machine with Intel® Core™ i7-6700HQ CPU @ 2.60GHz (only 1 core used) approximately 30 minutes for one dataset with ~20 000 ice leads and approximately 3 minutes for a dataset with ~ 8000 nodes. Depending on the number of used CPU cores, the parallelized version is several factors faster but needs a RAM of ~60GB for a dataset with 20,000 edges. In contrast, the non-parallelized version has constant RAM usage. In summary, two different experimental setups for graph translation are provided to match the different resources potential users have at hand. In practice, we run the code on a compute cluster instance with 64 GB memory and 8-CPUs (2x AMD EPYC Zen) in a parallelized manner.

For the network analysis, methods from the NetworkX package [5] proved useful. The properties of the graphs as well as the static graph link prediction method all stem from NetworkX. EvolveGCN was implemented with the help of the Pytorch Geometric Temporal package [18].

7 RESULTS AND DISCUSSION

This section presents the results obtained for the ice lead to graph translation, the spatio-temporal ice lead analysis, and the network analysis. The results are discussed here as well with a special focus on the limitations of the dataset and the chosen methods.

7.1 Ice Lead to Graph Translation

The ice leads could successfully be translated into NetworkX graphs. The addition of intersections is time-consuming but can be done. Subsequent analyses of the networks were possible and revealing. Per every single graph, few intersections (<100) were found compared to the respective large amount of edges (10k - 40k).

The biggest limitation of the provided dataset is that it does not yet contain any node summary. While easy to implement, running the algorithm to find close-by nodes is once again very time-consuming. We expect that the resulting network produced by employing node summaries would have much higher connectivity and higher degrees. See Figure 2 to get an intuition of the impact of node summaries: In the left corner, a zoomed-in picture can be seen that shows the detected ice leads. It becomes apparent that many lines are only shortly disrupted by ice - most likely caused by ice drift and ridges - and could be joined together to form fewer, yet longer ice leads. Indeed from farther away (zoomed picture on the right), the underlying ice lead network can be seen, since the disruptive parts are not visible anymore from distance. Node summary will be an essential feature for capturing the parts of the ice lead networks that become hidden under shallow ice debris.

7.2 Ice Lead Analysis

A primary focus of this work is to understand the temporal and geographic shifts these ice lead networks experience. In the following the geographical distribution, as well as temporal distribution of width, length, size, and total area of the ice lead networks are presented and discussed.

In Figure 3 the geographical distribution of ice leads throughout one year can be seen. There are clear differences - especially in the number of ice leads - over the four months which can be attributed to the seasonal cycles of the Arctic. Interestingly, the ice leads wander closer to the center of the Arctic during the winter months, while at the same time, expanding further south. The south expansion in winter is simply caused by a larger ice cover that is

[5] Code will be published here: https://github.com/liellnima/iceagle

not present in the summertime. The fact that ice leads are mostly occurring at the edges of the arctic sea-ice during the summertime, is a phenomenon that remains to be examined further.

We also looked at the geographical distribution of ice leads on the same day over a nearly 20-year span (see Figure 13 in the appendix). Clearly, these networks look very different year-to-year and have experienced quite drastic shifts over the past two decades. These shifts could, however, be simply attributed to the different ice sheet dynamics that produce different distributions of ice leads each year. Still, one dense cluster of ice leads that appears at the left corner of the images first decreases and then disappears completely in the end. A large ice floe that broke off the Arctic due to global warming and climate change could be an explanation for this observation.

The size of the total ice area of the ice lead networks goes through seasonal shifts, as can be seen in Figure 5 and Figure 7. Furthermore, both show a slight decrease over the complete time span (see Figure 6 for the size). The seasonal shifts can be attributed to the ice loss during the summertime and we suspect this is due to the increased surface area of water that comes from melting. It is important to mention that there are very few data points during the summer months because of the intense cloud coverage over the Arctic during this period. Still, the seasonal trend becomes apparent. The overall trend of decreasing ice lead numbers in Figure 6 can similarly be attributed to the loss of average annual Arctic sea-ice in the last years.

We were curious to see how the average width of ice leads varies over time, given the width of an ice lead determines the amount of heat loss it may experience from the ocean into the atmosphere, accelerating melting processes in the Arctic [11]. These results, presented in Figure 4 show seasonal trends (rather small as can be seen in Figure 8 in the appendix). However, they lack a strong long-term structure and pattern beyond a gentle negative slope in the linear regression fit. The gentle negative slope indicates that the ice leads are decreasing in width. Thinner ice leads actually cause greater heat fluxes, indicating that a positive feedback loop may be at play. Notably, there are some extreme outliers observed. Further work in identifying the root causes of these outliers is recommended to complement the development of prediction methods in this space, to ensure that tail events are adequately captured.

The length of the ice leads encodes to a certain degree how stable the ice floes are - a high number of long ice leads indicates that more ice floes are fragmented and are losing their integrity. The seasonal trends of the ice lead lengths are not highly obvious (see Figure 9 in the appendix), but similarly to the width graphs, a high number of outliers can be observed in the overall time trend (see Figure 10. There is a recent accumulation of outliers in the last years (few very long ice leads), indicating that the integrity of some ice floes might be increasingly endangered more recently.

While it is very well known that ice leads are subjected to seasonal trends and that ice sheets are constantly shifting, some of the described results are actually new: The trend that the number and of ice leads is decreasing is new and fits into the observation of shrinking ice coverage. The slight decrease in width is revealed here for the first time and should be investigated further due to the possibility of discovering a positive feedback loop here. The anomalies of ice lead lengths in recent years is another new discovery from the ice lead analysis that could be used for the analysis of ice floe stability.

7.3 Network Analysis

The network analysis showed surprisingly that the networks presented here are considerably different from common real-world networks. Due to their unique characteristics, common network science methods are not able to perform link prediction and graph classification on this type of network. We highlight, that the networks carry information that could be used to tackle these tasks, however, they are encoded in different properties than common real-world networks.

7.3.1 Network Properties. While common real-world networks are highly connected graphs, with multiple motifs and high transitivity, subject to the small-world phenomenon and shrinking diameters over time, the ice lead networks exhibit none of the properties. The networks exhibit the following properties:

- Planar
- Euclidean
- Very short paths (mostly node pairs present)
- No motifs such as triangles
- No transitivity
- Disconnected, connected components close to total number of edges
- No cluster or communities
- No small-world phenomenon
- Diameter follows seasonal trends
- Linear degree distribution on a log scale (and not log-log scale, see Figure 12)

We did not expect to find these networks diverge so strongly from common most real-world networks and want to highlight that ice leads pose an interesting, new and unexplored network class that calls for further investigation.

Some might argue that such networks seem to lack relevant interactive information that is usually encoded in these properties. However, the ice lead networks clearly do not lack structure or identifiable characteristics, it is simply encoded through different properties of the network than other real-world networks. The structure of the network for example is not defined via its connections, but much more via its positions in the euclidean space. Furthermore, no clusters or communities can be detected in these networks, however, when examining a picture of the networks as given in Figure 2, it becomes apparent that there are geographical clusters in the graphs. They may not be encoded in the connections of the network, but rather by their proximity and similarity (curves following the same paths and patterns, i.e. this information is encoded in the euclidean space). No motifs can be found, but different and repeating patterns can be seen nonetheless.

7.3.2 Network Methods. All the link prediction methods that we applied to the ice lead dataset failed - we suspect that the main reason for that is the unique characteristics described in the previous section.

The link prediction methods for the static graphs fail no matter how much percent (5 %, 10%, 20%) of the edges are dropped. All methods fail to predict even one missing edge correctly, i.e. that

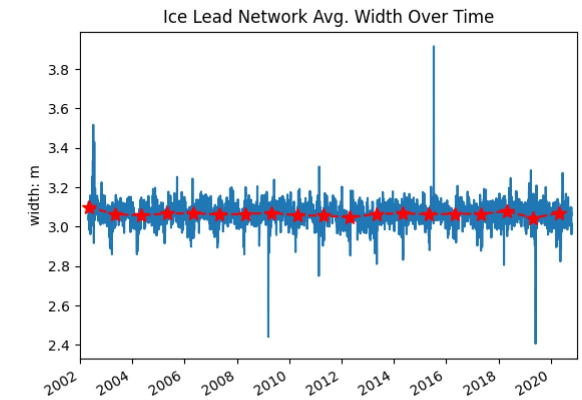

Figure 4: The average width of ice leads over time. The average of each year is included as red stars. Slope of annual average regression line: $-5.49 \cdot 10^{-4}$ meters per year.

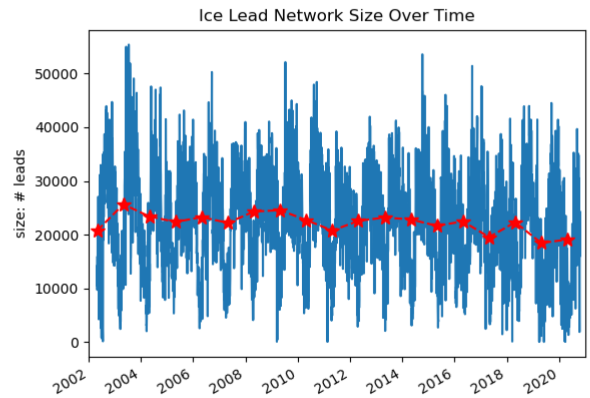

Figure 6: The size of ice lead networks over time. The average of each year is included as red stars. Slope of annual average regression line: -193.9 ice leads per year.

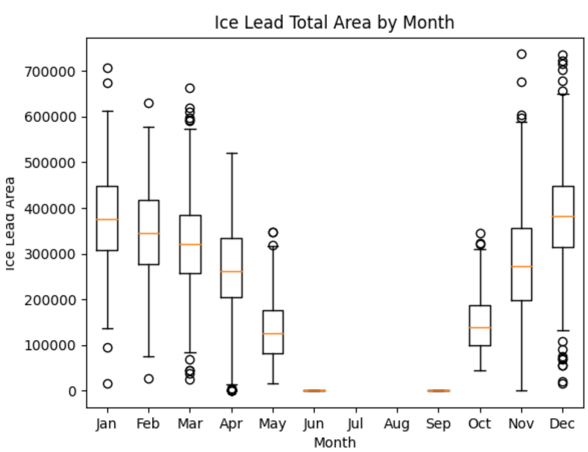

Figure 5: Total ice lead area by month across the entire dataset.

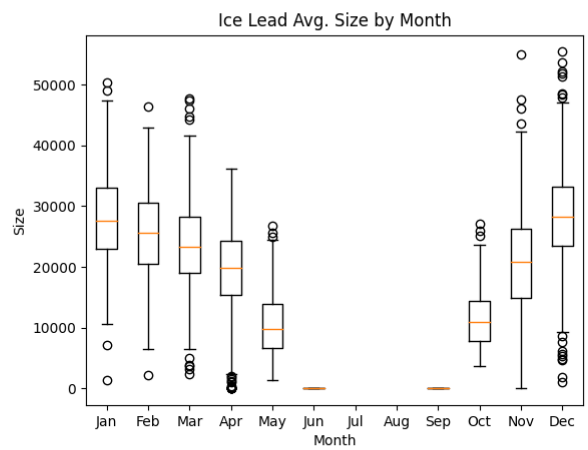

Figure 7: Network size (number of ice leads) by month across the entire dataset.

accuracy values and other metrics do not exceed 0.00 significantly. Table 1 lists our reasoning why each link prediction method failed. The Jaccard Coefficient increases with the number of common neighbours between two nodes - with an average degree of 1 there are practically no common neighbours and the method can only fail. The Adamic Adar Index calculates similarity scores based on shared features and weights the scores according to degree. Once again the common neighbours, and additionally the euclidean nature of the node attributes make it practically impossible to predict new links. The Resource allocation index fails for very similar reasons, since it works similarly like Adamic Adar Index, and only weights the scores differently. Preferential Attachement works especially well for scale-free networks, where large degree nodes are densely connected and low degree nodes are rarely connected. Since the ice lead network is sparsely connected and does not follow scale-free-laws, preferential attachment cannot perform well.

The main problem of the static link prediction methods might be that they are all from the domain of local similarity indices. Global and quasi-local similarity indices should be tried as well since they consider the whole topography of the network. Since the overall ice lead topography is very homogeneous and does not reveal new information, the authors are sceptical if those methods would perform better, however for completeness sake they could be considered in future work. To generate a complete picture of static link prediction, additionally dimensionality reduction-based and probabilistic/maximum-likelihood-based might be of interest. [9] We would like to highlight, that from all link prediction methods node embeddings (from the domain of dimensionality reduction-based methods) might be the most promising ones, since the euclidean information could theoretically be encoded in the embeddings. However, even the node embedding method must be adapted, since the underlying similarity measurements are currently based on network topography. To summarize, all alternative static link prediction methods rely on topographic similarity, which the ice lead networks cannot provide.

The main reason why all these methods fail - and also further common link prediction methods will fail - is that the relevant information exploited from link prediction methods is the structure of the networks. As can be seen in the zoomed-in figure in Figure 2, the ice lead graph consists mostly of single edges that are not connected with each other. Since scaling the edge-dropping (e.g. to 20%) does not affect the performance, we conclude that necessary similarity information is not only *less* encoded into the topographic, but rather *not encoded at all*. There are two paths to solve this general problem: (1) A link prediction method is applied that does not exploit the structure of a network to derive new linkages, but rather its euclidean structure. (2) The dataset is adapted such that relevant information is included in the topography of the network and not in the node attributes / euclidean space.

Also the dynamic link prediction method (learning-based) - the EvolveGCN - failed on the time-series link prediction task. The implementation of the EvolveGCN algorithm already posed many obstacles, since e.g. no edge variation is expected within the time-series. While of all learning-based link prediction methods EvolveGCN allows the largest variability of graphs, the ice lead networks still demand more variability. The set of nodes, its node attributes, the set of edges, and its edge attributes - all of these are changing from one time step to another. While the algorithm can be adapted to accept even this amount of variability, the method itself needs at least one scale that remains fixed to succeed. We conclude so far, that EvolveGCN can currently not perform link prediction on the ice lead networks.

There are no other alternative methods for dynamic link prediction available so far that could coop with the amount of variability present in the ice lead networks. In order to make dynamic link prediction possible two path could be followed: (1) Developing a new method similar to EvolveGCN with the euclidean space as a fixing point; (2) Fixing one scale in the dataset, e.g. by tracking the ice leads which would result in a fixed set of edges.

The graph classification with EvolveGCN faces the same challenges and problems as the link prediction task and fails similarly. However, the graph classification task can be simplified further to a pure static graph classification problem, which might overcome some of the problems.

We believe that link prediction and node classification are still possible on the network dataset, however, further research is necessary for this. The information in the euclidean space can be used and furthermore, tracking of the nodes could create a fixed set of edges that consequently also enables the use of EvolveGCN.

Methods	Mode	Reasoning
Jaccard Coefficient	Static	Not enough common neighbours
Resource Allocation Index	Static	Nearby nodes don't share common features
Adamic Adar Index	Static	Nearby nodes don't share common features
Preferential Attachement	Static	Sparsely connected nodes
EvolveGCN	Dynamic	High temporal variability

Table 1: The methods used for the analysis and the reasoning for their failures.

7.4 Future Work

Future work includes improving the ice lead network construction algorithm to incorporate node summarization in order to provide more realistic networks. Furthermore, the results from the networks and ice lead analysis can be improved via more thorough collaboration with cryospheric experts to help uncover the meaning behind these findings for the characterization of ice leads and their dynamics. The current link prediction and graph classification methods in the field of network science need to be modified to be able to perform adequately on ice lead networks. We believe that EvolveGCN could be adapted and transformed to handle graphs that have a higher spatial variability and range in the euclidean space. Further, we propose that looking into the task of ice lead tracking from a network science perspective; i.e. Node embeddings and alignment might be the path forward towards ice lead tracking more efficiently and accurately. In summary, looking at ice leads from a network science perspective opens many new doors and could be used to bring actionable insights to the climate science community as well as serve as a case study for the machine learning community in creating more generalizable network science algorithms and methods.

8 CONCLUSION

In this project, we discovered that ice leads can be analyzed from a network science perspective and that the ice lead graph properties diverge from common characteristics of real-world networks. As a result, the application of current link prediction techniques failed on these graphs. This is not to say ice lead networks lack structure, as they exhibit identifiable unique features and characteristics that could be leveraged in novel forecasting methods.

The successful ice lead analysis confirmed that ice leads are subject to seasonal changes and that their width is power-law distributed (see also [11]). It additionally revealed more global trends, such as that the size of ice lead networks is shrinking, the width of ice leads is slightly decreasing and the length distribution shows more anomalies in recent years. The authors offer shrinking sea-ice coverage as a possible explanation of these phenomena, however, it remains open to attribute these changes to a concrete cause.

We believe that a network science perspective on ice lead networks is meaningful to both the cryospheric community as well as the network science community. Novel cryospheric insights, e.g. regarding ice floe stability, can be drawn from this perspective, while the unique structure of ice lead networks makes them a fantastic case study for the development of more generalizable algorithms. The network data will be made readily available [6] and we recommend machine learning practitioners to collaborate with domain experts to build off of this work.

Ice leads play a key role in Arctic heat fluxes [11] and can be used to predict sea ice loss [29]. Consequently, improving tracking, analysis, and forecasting of ice leads is essential to enhance our understanding of the Arctic sea ice and climate system. The value of this knowledge becomes clear when looking at the effects of sea ice loss: Millions of people must relocate, many wildlife species lose their habitat, methane release is increased, and more [27]. This project contributes to understanding the Arctic processes that we need to fathom to mitigate climate change and adapt to its consequences.

ACKNOWLEDGMENTS

We would like to thank Prof. Dr. Gunnar Spreen who provided us with insights into ice lead detection and possible challenges connected to the project idea.

REFERENCES

[1] José S Andrade Jr, Hans J Herrmann, Roberto FS Andrade, and Luciano R Da Silva. 2005. Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs. *Physical review letters* 94, 1 (2005), 018702.

[2] David Bröhan and Lars Kaleschke. 2014. A nine-year climatology of Arctic sea ice lead orientation and frequency from AMSR-E. *Remote Sensing* 6, 2 (2014), 1451–1475.

[3] Lucia Carlucci, Gianfranco Ciani, and Davide M Proserpio. 2003. Borromean links and other non-conventional links in 'polycatenated'coordination polymers: re-examination of some puzzling networks. *CrystEngComm* 5, 47 (2003), 269–279.

[4] M. Dirnberger, T. Kehl, and A. Neumann. 2015. NEFI: Network Extraction From Images. *Scientific Reports* 5 (2015). http://dx.doi.org/10.1038/srep15669

[5] Aric Hagberg, Pieter Swart, and Daniel S Chult. 2008. *Exploring network structure, dynamics, and function using NetworkX.* Technical Report. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

[6] Jay P Hoffman, Steven A Ackerman, Yinghui Liu, and Jeffrey R Key. 2019. The detection and characterization of Arctic sea ice leads with satellite imagers. *Remote Sensing* 11, 5 (2019), 521.

[7] Nils Hutter, Lorenzo Zampieri, and Martin Losch. 2019. Leads and ridges in Arctic sea ice from RGPS data and a new tracking algorithm. *The Cryosphere* 13, 2 (2019), 627–645.

[8] Jon Kleinberg. 2000. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing.* 163–170.

[9] Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, and Bhaskar Biswas. 2020. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications* 553 (2020), 124289. https://doi.org/10.1016/j.physa.2020.124289

[10] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining.* 177–187.

[11] S Marcq and J Weiss. 2012. Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere. *The Cryosphere* 6, 1 (2012), 143–156.

[12] Thorsten Markus, Donald J Cavalieri, and Alvaro Ivanoff. 2011. Algorithm theoretical basis document: Sea ice products: Updated December 2011.

[13] Dmitrii Murashkin and Gunnar Spreen. 2019. Sea Ice Leads Detected From Sentinel-1 SAR Images. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium.* IEEE, 174–177.

[14] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. 2020. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5363–5370.

[15] Pierre Rampal, Sylvain Bouillon, Einar Ólason, and Mathieu Morlighem. 2016. neXtSIM: a new Lagrangian sea ice model. *The Cryosphere* 10, 3 (2016), 1055–1073.

[16] Fabian Reiser, Sascha Willmes, and Günther Heinemann. 2020. Daily sea ice lead data for Arctic and Antarctic. https://doi.org/10.1594/PANGAEA.917588

[17] Fabian Reiser, Sascha Willmes, and Günther Heinemann. 2020. A new algorithm for daily Sea Ice lead identification in the arctic and antarctic winter from thermal-infrared satellite imagery. *Remote Sensing* 12, 12 (2020), 1957.

[18] Benedek Rozemberczki, Paul Scherer, Yixuan He, George Panagopoulos, Alexander Riedel, Maria Astefanoaei, Oliver Kiss, Ferenc Beres, , Guzman Lopez, Nicolas Collignon, and Rik Sarkar. 2021. PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management.* 4564–4573.

[19] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. 2020. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the 13th International Conference on Web Search and Data Mining.* 519–527.

[20] RC Schnell, Roger Graham Barry, MW Miles, EL Andreas, LF Radke, CA Brock, MP McCormick, and JL Moore. 1989. Lidar detection of leads in Arctic sea ice. *Nature* 339, 6225 (1989), 530–532.

[21] Dexuan Sha, Younghyun Koo, Xin Miao, Anusha Srirenganathan, Hai Lan, Shorojit Biswas, Qian Liu, Alberto M Mestas-Nuñez, Hongjie Xie, and Chaowei Yang. 2021. Spatiotemporal Analysis of Sea Ice Leads in the Arctic Ocean Retrieved from IceBridge Laxon Line Data 2012–2018. *Remote Sensing* 13, 20 (2021), 4177.

[22] Joakim Skardinga, Bogdan Gabrys, and Katarzyna Musial. 2021. Foundations and modelling of dynamic networks using dynamic graph neural networks: A survey. *IEEE Access* (2021).

[23] National Snow and Ice Data Center. 2021. Sea Ice Features : Leads. In All About Sea Ice. Accessed 18 October 2021. https://nsidc.org/cryosphere/seaice/characteristics/leads.html.

[24] EOS Project Science Office The Earth Observatory. 2013. Extensive Ice Fractures in the Beaufort Sea. Accessed 12 November 2021. https://eoimages.gsfc.nasa.gov/images/imagerecords/80000/80752/beaufort_vir_2013054_lrg.jpg.

[25] Casper Van der Wel. 2019. Introducing PyGEOS. Accessed 12 November 2021. https://caspervdw.github.io/Introducing-Pygeos/.

[26] Timo Vihma. 2014. Effects of Arctic sea ice decline on weather and climate: A review. *Surveys in Geophysics* 35, 5 (2014), 1175–1214.

[27] John Walsh. 2013. Melting Ice: What is Happening to Arctic Sea Ice, and What Does It Mean for Us? *Oceanography* 26 (06 2013). https://doi.org/10.5670/oceanog.2013.19

[28] Andreas Wernecke and Lars Kaleschke. 2015. Lead detection in Arctic sea ice from CryoSat-2: quality assessment, lead area fraction and width distribution. *The Cryosphere* 9, 5 (2015), 1955–1968.

[29] Yuanyuan Zhang, Xiao Cheng, Jiping Liu, and Fengming Hui. 2018. The potential of sea ice leads as a predictor for summer Arctic sea ice extent. *The Cryosphere* 12, 12 (2018), 3747–3757.

[30] Fan Zhou, Xovee Xu, Ce Li, Goce Trajcevski, Ting Zhong, and Kunpeng Zhang. 2020. A heterogeneous dynamical graph neural networks approach to quantify scientific impact. *arXiv preprint arXiv:2003.12042* (2020).

[6] After consultation with the authors of the raw ice lead dataset [6]

9 APPENDIX

Supplementary figures for additional results.

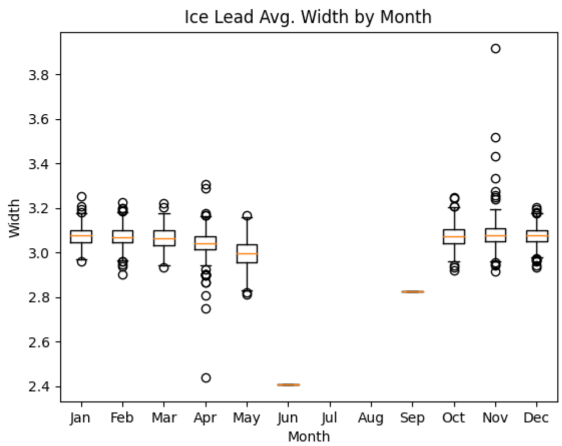

Figure 8: Ice lead width by month.

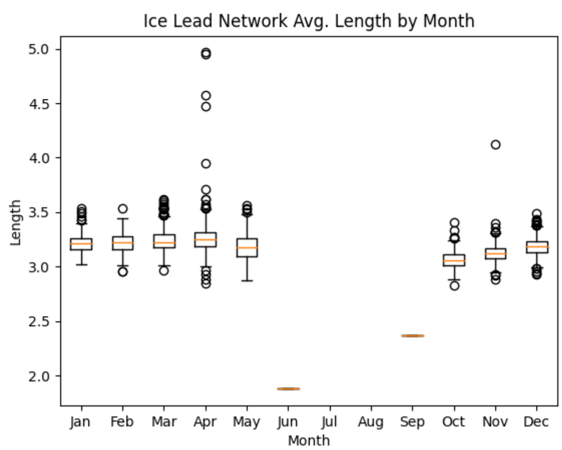

Figure 9: Ice lead length by month.

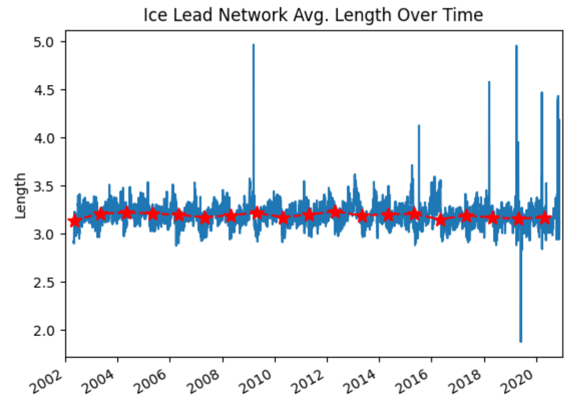

Figure 10: Ice lead average length over time.

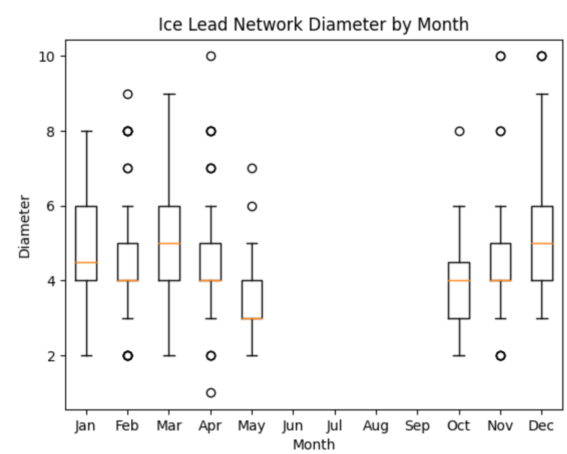

Figure 11: Ice lead network diameter by month.

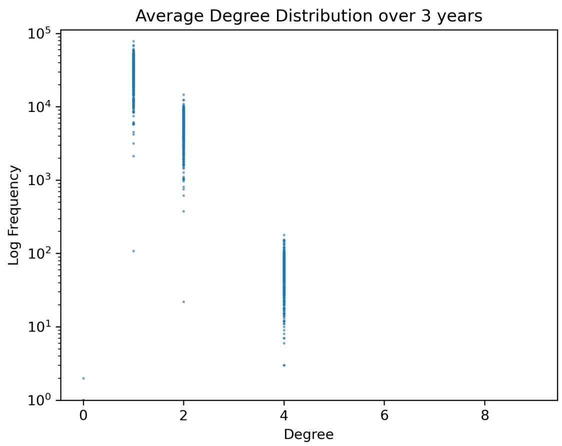

Figure 12: Degree distribution of ice leads over three years.

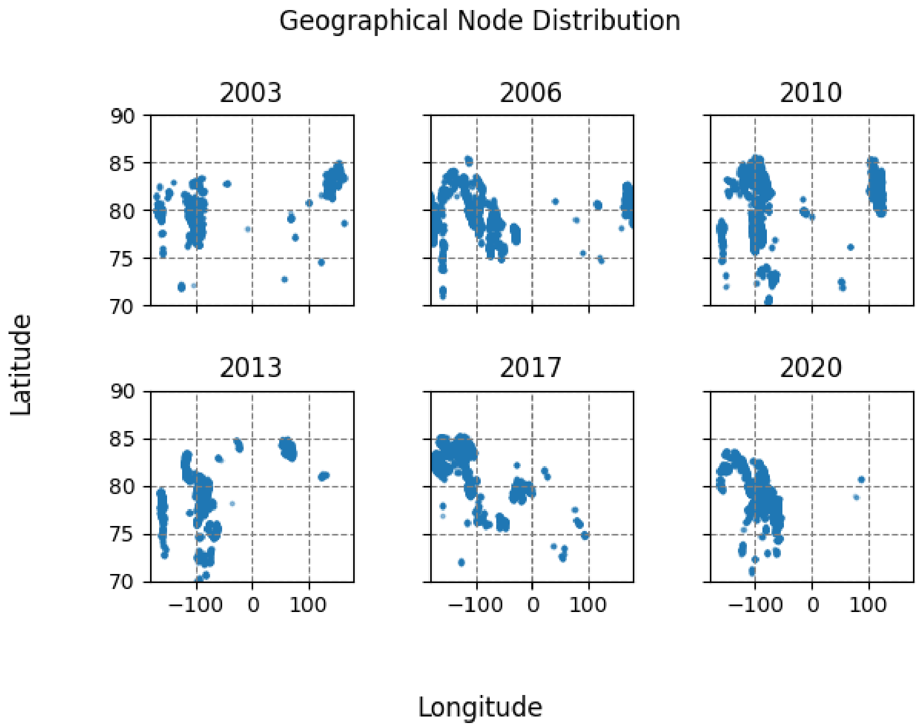

Figure 13: The geographical distribution of ice leads over the span of the entire dataset (2002 - 2020).